Applied Soft Computing 70 (2018) 423–464

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Review Article

Soft Computing based object detection and tracking approaches: State-of-the-Art survey

Manisha Kaushal[a], Baljit S. Khehra[b], Akashdeep Sharma[a,*]

[a] UIET, Panjab University, Chandigarh, India
[b] Computer Science Engineering, BBSBEC, FatehgarhSahib, Punjab, India

ARTICLE INFO

Article history:
Received 8 July 2016
Received in revised form 14 March 2018
Accepted 16 May 2018
Available online 26 May 2018

Keywords:
Soft computing
Object detection and tracking
Neural network and deep learning
Fuzzy logic
Evolutionary algorithms

ABSTRACT

In recent years, analysis and interpretation of video sequences to detect and track objects of interest had become an active research field in computer vision and image processing. Detection and tracking includes extraction of moving object from frames and continuous tracking it thereafter forming persistent object trajectories over time. There are some really smart techniques proposed by researchers for efficient and robust detection or tracking of objects in videos. A comprehensive coverage of such innovative techniques for which solutions have been motivated by theories of soft computing approaches is proposed. The main objective of this research investigation is to study and highlight efforts of researchers who had conducted some brilliant work on soft computing based detection and tracking approaches in video sequence. The study is novel as it traces rise of soft computing methods in field of object detection and tracking in videos which has been neglected over the years. The survey is compilation of studies on neural network, deep learning, fuzzy logic, evolutionary algorithms, hybrid and recent innovative approaches that have been applied to field of detection and tracking. The paper also highlights benchmark datasets available to researchers for experimentation and validation of their own algorithms. Major research challenges in the field of detection and tracking along with some recommendations are also provided. The paper provides number of analyses to guide future directions of research and advocates for more applications of soft computing approaches for object detection and tracking approaches in videos. The paper is targeted at young researchers who will like to see it as platform for introduction to a mature and relatively complex field. The study will be helpful in appropriate use of an existing method for systematically designing a new approach or improving performance of existing approaches.

© 2018 Elsevier B.V. All rights reserved.

Contents

1. Introduction 424
2. Motivation and contribution 426
3. Neural network based object detection and tracking approaches 427
   3.1. Neural network based object detection 427
   3.2. Neural network based object tracking 430
   3.3. Discussion 432
4. Fuzzy logic based approaches 434
   4.1. Fuzzy logic based detection approaches 434
   4.2. Fuzzy logic based object tracking approaches 437
   4.3. Discussion 439
5. Evolutionary algorithm based approaches 439
   5.1. Evolutionary algorithm based object detections 439
   5.2. Evolutionary algorithm based object tracking 443
   5.3. Discussion 445

* Corresponding author.
E-mail addresses: er.manishakaushal@gmail.com (M. Kaushal), baljit.singh@bbsbec.ac.in (B.S. Khehra), akashdeep@pu.ac.in (A. Sharma).

https://doi.org/10.1016/j.asoc.2018.05.023
1568-4946/© 2018 Elsevier B.V. All rights reserved.

6. Hybrid approaches 445
6.1. Discussion 448
7. Other trends in object detection and tracking 449
7.1. Learning based systems 449
7.2. Knowledge based system 450
7.3. Expert system based approaches 450
7.4. Tracking-by-detection 451
7.5. Ensemble tracking 451
7.6. Correlation filters based studies 452
8. Datasets 452
9. Challenges in object detection and tracking in videos 455
10. Analysis and inferences 456
10.1. Selection of algorithms: some guidelines 459
11. Conclusions and summarizations 460
References 460

1. Introduction

The exponential growth in hardware facilities like cameras, processing machines, mobile phones have led to an explosion of studies in automated video analysis for object detection and tracking. It is one of the hottest topics of research in computer vision and image processing. Object detection and tracking in video sequence is the key technology in the development of various video analysis applications that tires to detect and track objects over a sequence of images by replacing old traditional methods of monitoring cameras by human operators. The proposed solutions range from low cost handheld devices or cameras to high cost sophisticated and proprietary solutions. Object detection is the process of locating the occurrence of object using number of techniques like background subtraction, feature extraction, statistical methods etc. Fig. 1 shows some example of object detection in videos. Colored boxes on the figures are used to highlight the detected object on a frame. A number of traditional approaches like background detection, frame difference, Gaussian mixture modeling, optical flow based techniques are commonly used to detect objects. Object tracking, on the other hand is process of monitoring an object or multiple objects using a camera over time. Tracking deals with assigning labels to object being tracked and plotting suitable trajectories to specify object motion or deviations. Addition object specific information like area, shape size, orientation can also be extracted by object tracking method. Tracking methods work by detecting the object when it appears for the first time in a frame and predict its trajectories. Such detection-based algorithms estimate the object location in every frame independently. These require an offline training stage and cannot be applied to unknown objects. There is another class of methods where detectors and trackers work simultaneously in which tracker can provide weakly labeled training data for a detector and can help to improve its performance using suitable learning mechanisms. Such trackers can work in online as well offline mode. Fig. 2 shows sample tracked objects over sequence of frames.

The general framework for object detection and tracking in videos is shown in Fig. 3. Video images or frames are processed to detect object using an algorithm which is followed by labeling and tracking of that object. Labeling is important to differentiate and track an object over series of frames. Preprocessing operations can be applied before or after an object is detected. Preprocessing before object extraction will target to enhance quality of video sequences by suppression of noise or redundant information to enable more accurate and precise detections. It is different from post processing where target is to enhance boundaries or shapes of detected object with help of certain morphological operations. Object detection and tracking go hand in hand in which detected object is tracked in subsequent frames. Detection and tracking are preliminary steps towards developing various applications like vehicle extraction from traffic surveillance scenes [1,2], detection of suspicious behaviors in shopping malls in real time [3], path detection in video surveillance [4], observation of birds intelligently[5], cyclist motion detection [6] and can be further utilized in several application such as aviation protection and safety.

Detection and tracking techniques have flooded in the recent past but number of challenges involved in this process makes this problem an evergreen research problem. Challenges that complicate detection and tracking problems include camera jitter, rapid illumination changes, dynamic backgrounds, moving cameras and shadow detection etc. These problems cannot be solved by simple algorithms because of complexities, impreciseness and improbable factors introduced at various intermediate steps. To cater these problems soft computing based studies had begun to take center stage as soft computing techniques have demonstrated a sound history in tackling these issues which are difficult to be handled by traditional approaches. In Fig. 4, the evolution of soft computing based approaches for object detection and tracking in videos in terms of number of papers published over last 17 years has been portrayed. The growth in early years was bit slower but it has gradually been exponential in the past 6–7 years. The slope of curve indicates a rapid rise in number of publications in previous years from 2011 onwards because these approaches have established themselves as most active areas of research in field of object detection and tracking. Soft computing based techniques have demonstrated tremendous success in handling complex applications and because of their dazzling abilities, these techniques have been able to find acceptability among masses. These reasons have made soft computing based object detection and tracking approaches popular for solving various real time applications such as visual surveillance, analysis of sports, human action recognition, event detection and so on. A number of soft computing based object detection and tracking approaches in video sequence have been worked upon in the recent past by various researchers which have demonstrated promising results. This paper is a novel attempt to cite the efforts put by those researchers in this ever growing field.

This manuscript has been structured under two major headings: object detection and tracking. Subcategories of Neural Networks, Fuzzy Logic, Evolutionary and Hybrid Approaches are framed under these headings. Studies under these subheadings were harvested using a systematic process. Some newly explored studies based on representation learning, knowledge based studies, ensemble tracking have also been included. An extensive set of studies from conferences and research journals was prepared at first level in which soft computing techniques have been used by researchers. This list was sorted and refined further to include studies in which soft computing techniques have played major and noticeable role in object detection or tracking. An attempt was made to make the study exhaustive however recent studies and from reputed

Fig. 1. Examples of object detection in videos.

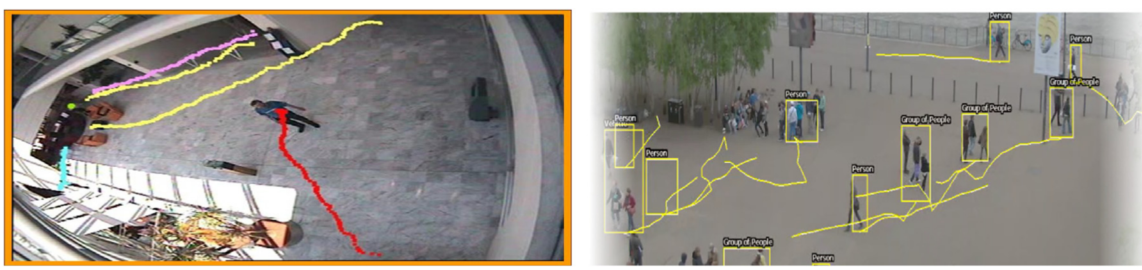

Fig. 2. Example of object tracking in videos [129].

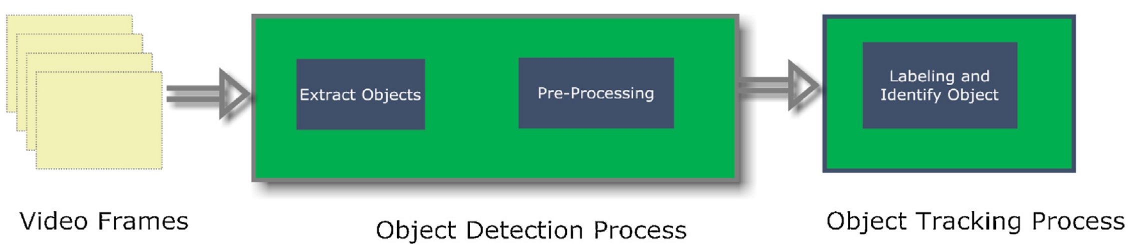

Fig. 3. General framework of object detection and tracking in videos.

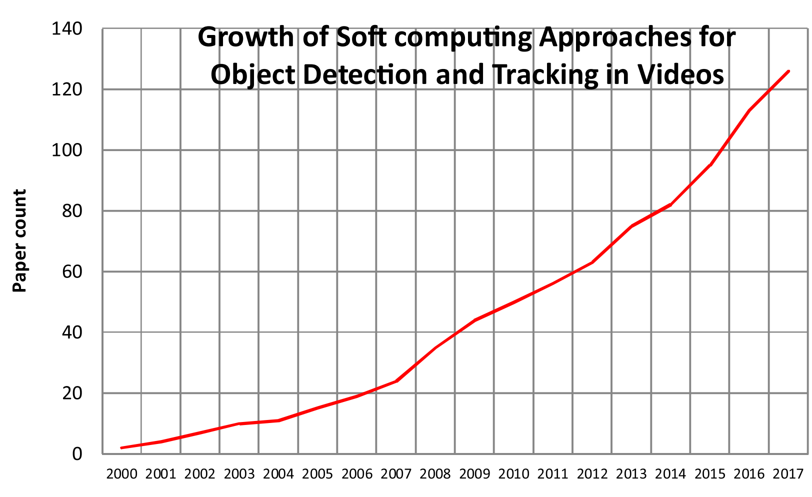

Fig. 4. The graph shows cumulative number of relevant publication and their span among the years 2000–2017 related to Soft computing based approaches for object detection and tracking in videos.

Table 1
Summarization of various Survey papers on Object detection and tracking in videos.

Title	Author	Description	Year	Paper
Background subtraction techniques: A review	M. Piccardi	Provides comprehensive survey on various Background subtraction techniques and also compares techniques in terms of speed, memory requirements and accuracy.	2004	7
Object Tracking: A Survey	A. Yilmaz, O. Javed, M. Shah	Presented a classification of tracking methods with their pros and cons. Important issues such as selection of motion models, and detection of objects, use of appropriate image features were also discussed.	2006	8
Background modeling using mixture of Gaussians for foreground detection: A survey	T. Bouwmans, F. E. Baf, B. Vachon	Overview of traditional Background modeling approach using Gaussian mixture model for foreground detection were reviewed; different improvements w.r.t. traditional approaches found in the literature were categorized.	2008	9
Review and Evaluation of well-known methods for moving object detection and tracking in videos	B. Karasulu	A systematic analysis of various known object detection and tracking algorithms with focus of various performance measures. Various metrics have been proposed and effectiveness of algorithms is computed via various metrics.	2010	10
Recent advanced statistical background modeling for foreground detection: A systematic survey	T. Bouwmans	Surveyed recent developments in statistical background modeling approaches for foreground detection.	2011	11
Recent advances and trends in visual tracking: A review	H. Yang, L. Shao, F. Zheng, L. Wang, Z.Song	Recent progress of various state-of-the art feature descriptors, online learning methods and context information methods was presented with positives and negatives.	2011	193
Systematic Survey on Object Tracking Methods in Video.	J. J. Athanesious, P. Suresh	Analysis of various object tracking methods in video are reported and reviewed.	2012	12
Video Surveillance Systems – A Survey.	C. L. Devasena, R. Revathí, M. Hemalatha	Presented a review of the state-of-the-art methods with a focus on video surveillance only	2011	13
A Survey on Moving Object Detection and Tracking in Video Surveillance System	K. A Joshi, D. G. Thakore		2012	14
Detection of Objects in Motion-A Survey of Video Surveillance	J. Raiyn		2013	15
Study of Moving Object Detection and Tracking for Video Surveillance.	A. K. Chauhan, D. Kumar		2013	16
Automated human behavior analysis from surveillance videos: a survey.	D. Gowsikhaa, S. Abirami, R. Baskaran	Reviewed few algorithms in the domain of human behavior analysis which includes both low level and high level techniques with focus on semantically enhanced analysis.	2014	17
A Comprehensive Review Of Background Subtraction Algorithms Evaluated with Synthetic and Real Videos	A. Sobral, A. Vacavant	Overview of traditional background subtraction approach in videos is presented and compared a wide range of Background Subtraction algorithms from literature.	2014	18
Survey on Background Modeling and Foreground Detection for Real Time Video Surveillance.	S. Jeevaa, M. Siva balakrishnan	Presented a review and comparative study on Background modeling and Foreground detection for real time video surveillance.	2015	19
Fuzzy human motion analysis: A review	C. H. Lim, E. Vats, C. S. Chan	Presented significant contributions published in literature that use Fuzzy oriented human motion analysis approaches and explored various future research directions.	2015	20
Human detection from images and videos: A survey.	D. T. Nguyena, W. Lia, P. O. Ogunbona	A review on the state-of-the-art for human detection techniques is presented. Challenges such as occlusion and real-time human detection along with open issues and future research directions are highlighted.	2016	21

databases are included for obvious reasons of quality and wider dissemination. At next level of manuscript preparation, some studies with good number of citations have been added under some subheadings where appropriate number of studies was not available. The audiences of this paper include young practitioners and researchers who see this domain as wider research challenge. The paper has been targeted at young researchers who are motivated by powers and success of soft computing approaches to handle complex real time environments or situations. The study is novel as the authors were not able to find a survey that has provided a comprehensive compilation of work on soft computing approaches only. There have been some very good survey papers in recent past as shown in Table 1 but none of them focuses on the specified topic as a whole. The paper will be beneficial to readers around the globe who are looking for exploration in this field.

2. Motivation and contribution

Soft Computing comprises of number of methodologies such as Neural network, Genetic Algorithms, Fuzzy logic, evolutionary approaches aimed at exploiting tolerance for uncertainty and imprecision, to achieve low cost and robust solutions for real world applications. The human mind is a role model for soft computing based approaches and nature is the driving force behind it. Theories based on nature together with capabilities of human mind had proposed some revolutionary solutions in the field of object detections and tracking in videos. This study is aimed at reviewing such state of art methods on this topic which are very difficult to find. The work presented in this paper is intended to provide a wide overview on use of various soft computing based approaches for object detection and tracking in videos. Our study is driven by the desire to identify various soft computing approaches for object detection and tracking in videos and group most of them together for benefit of research community as a whole. To the best of our knowledge, most surveys research papers focus only on well known moving object detection and tracking approaches targeted at specific applications. There has been little mention of soft computing approaches in these surveys moreover these studies have been projected as supporting theories or as contributor towards specific applications. The contributions have not been fully acknowledged and this discussion is also randomly scattered. There has not been any survey to pit wonderful contributions that soft computing based approaches have been able to make in field of object detection and tracking in videos. The current study is in sharp contrast to past survey listed in Table1 as it has brought together studies on all soft computing approaches under one umbrella. Table 1 is arranged in chronological order of publication and summarizes major contributions made by these studies; discussions on previous surveys and how current study deviates from these available methods is presented here.

The review presented by Piccardi [7] has focused on various background subtraction techniques only and had produced comparison of different background subtraction based approaches in terms of various parameters. Bouwans et al. [9] has targeted background modeling techniques using Gaussian mixture based studies. A review of moving object detection and tracking targeting video surveillance applications have been provided by [13–16]. Studies by Sobral et al. [18]; Jeevaa et al., [19] are also focused on background subtraction techniques for video surveillance applications. A discussion on various metrics for evaluation of moving object detection and tracking methods was presented by Karasulu [10]. The study evaluates various algorithms on the basis of these metrics. Summarization of various statistical methods for object detection has been done by Bouwmans [11]. Yilmaz et al. [8] had demonstrated broad overview of various object tracking approaches in videos and also discussed issues related to tracking such as selection of motion models and detection of objects, use of appropriate image features etc. Yang et al. [193] traces progress of various state-of-the art feature descriptors along with online learning methods used in various tracking mechanisms. A similar sort of analysis for various object tracking methods have been presented by Athanesious et al. [12]. Some surveys on human motion behavior detection like [17] are also available in literature. Lim et al. [20] has presented significant contributions based on fuzzy oriented human motion analysis that are published in the literature. A recent paper on human detection has been proposed by Nguyena et al. [21]. The paper has discussed various studies for human detection focusing on challenges like occlusions and real time human detection.

The studies discussed above are focused towards discussion of related studies on specific applications like video surveillance [14,15], human behavior [17,21], background modeling [9,11,18] etc. Although compilation of studies targeting certain suitable applications is important but contributions of soft computing techniques in developing these applications shall not be ignored. As work in these applications prospered, together with challenges thrown by real time environmental conditions like background cluttering, occlusions, shape, appearance changes of object, studies based on soft computing techniques had really evolved over past few years. Soft computing approaches are tolerant of uncertainty, partial truth, and approximation and imprecision and have sound history in offering good solutions in field of object detection and tracking. A survey based on contribution of soft computing based studies for object detection and tracking will widely contribute in that direction and is likely to be appreciated by fellow researchers. This paper is an attempt in that direction and will focus to track development, use and progress of soft computing based methods in the development of some good applications. This article's contribution has many folds. Firstly, the survey enumerates and highlight numerous soft computing based methods in the field of object detection and tracking in videos and categorizes them into neural network based approaches, fuzzy logic based approach, evolutionary approaches, hybrid and some novel approaches. This type of categorization helps to gain insight into various categories of soft computing based techniques in fairer way and helps in evaluating their suitability for numerous applications. This is first survey to the best of author's knowledge focusing only on soft computing based approaches for object detection and tracking in videos. Secondly, the work provides comprehensive study to readers/researchers to understand soft computing based concepts for object detection and tracking in videos. The paper will act as platform for integration of two relatively complex fields of object detection and soft computing approaches bundling together all soft computing approaches at one platform. We provide broader prospective of the state-of-the-art rather than focusing on specific application for object detection and tracking in videos. The paper can contribute to attract young talent towards explorations, integration and innovations in the field of object detection and soft computing. The paper can also act as guiding tool for reviewers and researchers for evaluating other state of art methods. The paper can also contribute in identifying soft computing based object detection and tracking approaches which are available for various user applications with relative rate of accuracy. Analysis, inferences and challenges for object detection and tracking in videos are presented to help in identifying latest trends and guide readers to identify those techniques which best suit their needs. Discussions regarding strengths and weaknesses of each soft computing approach to handle different challenges are also provided for benefit of research peers. The survey also highlights techniques towards which future research is shifting and directions for some new techniques like deep learning, hybridization of deep learning with evolutionary approaches are also listed.

The organization of paper is as follows: Section 3 describes most commonly used neural network based object detection and tracking approaches for video sequences and outlines deficiencies and advantages associated with these approaches. This section also includes studies on deep learning techniques. Section 4 describes some fuzzy logic based approaches for object detection and tracking in videos. Section 5 reviews some of the work on evolutionary algorithms based approaches. Hybrid approaches for object detection and tracking in videos are covered in Section 6. Some recent novel techniques based on representation learning, knowledge based system, learning by detection, ensemble tracking and studies based on correlation filters are discussed in Section 7. Section 8 discusses about available datasets for evaluation of algorithms, list of studies using those datasets have also been discussed. Various challenges in field of object detection and tracking are provided in Section 9. Section 10 provides a number of analyses with brief guidelines for selection and application of specific algorithm. Last section concluded whole study and lists some crucial outcomes for further exploration.

3. Neural network based object detection and tracking approaches

Neural networks are family of models inspired by working of human brain and nervous system. Neural network have the ability to learn from historic data and neural based algorithms are powerful and proven tools for providing useful and efficient solutions for classification problems. As a model for simulating learning function of human brains, they are able to perform classification of moving objects accurately when network is trained with sufficient training data. A substantial increase in number of neural networks approaches for object detection in videos has been observed in recent publications. With the advent of techniques like deep learning and deep convolution networks, success gained by these studies is also quite remarkable. The studies have been organized in hierarchy according to the neural architecture or algorithm employed in them. Studies are categorized and grouped according to proposed classification for better understanding and readability. Fig. 5 shows arrangement of available papers on neural network.

3.1. Neural network based object detection

An edge detection method followed by neural network based edge matching algorithm was developed by Dewan et al. [22]. The matching algorithm was based on integer distance transformation and feed forward neural network was used for classification of moving objects accurately. A combination of artificial neural network and Kalman Filter based approach for moving object detection captured by moving camera was proposed by Taleghani et al.

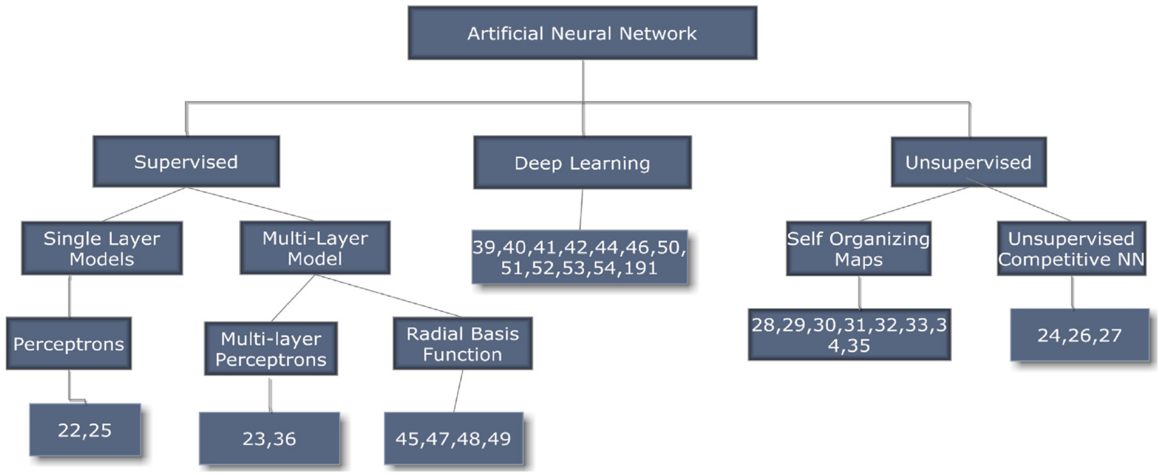

Fig. 5. Summary of neural network based object detection approaches.

[23]. In their work, videos were captured by moving camera and a multilayer perceptron was embedded in Kalman Filter for accurate detection of object motion. Unsupervised competitive neural network approach based on background subtraction for real time video segmentation was introduced by Luque et al. [24]. The neural architecture served both as an adaptive model for background subtraction and as classifier of pixels for foreground and background. As a result, efficiency was increased and there was reduction in processing time. Culibrk et al. [25] had put forward a neural based architecture to design a Bayesian classifier for video object segmentation. The proposed approach worked as an unsupervised classifier and did not require any training. The model was advantageous as it was capable of online learning. An improved background subtraction algorithm using unsupervised competitive neural network approach was proposed by Baena et al. [26]. In this case, synaptic weights and adaptive neighborhood of neurons acted as background model and were continuously refreshed to represent background statistics. The technique was pixel level parallelized and was targeted to achieve real time processing by hardware implementation. Another approach employing use of competitive neural network based on dipoles was presented by Baena et al. [27] to efficiently classify pixels as foreground or background. The approach worked by estimating the directionality of data for aid in effective detection of moving objects in videos. As a result, segmentation accuracy of proposed approach was improved as compared to Gaussian mixture and other Background Subtraction models. The drawback of approach was its failure to estimate moving object boundaries effectively. Self organizing maps with analysis of histogram were used by Baena et al. [28] for traffic flow monitoring. In their method, self-organizing neural network (SOM) was used for anomalous objects detection according to its shape. In order to handle occlusions efficiently, a procedure was applied to evaluate and identify objects integrated in an over-lapped zone. The overlapping of extracting objects was reduced after analyzing horizontal and vertical histograms of binary image. Zhou et al. [29] had proposed an automatic pathological diagnosis procedure which utilizes artificial neural network ensemble to identify lung cancer cells in the images of the specimens of needle biopsies. The approach uses two-level ensemble neural architecture composing of heterogeneous ensembles that not only comprises individual networks with different number of output units but also employs different methods to combine individual predictions. The first-level ensemble is utilized to judge whether a cell is a cancer cell with high confidence. The predictions of those individual networks were combined by a novel prediction method to improve the accuracy of false negative identification in the first-level ensemble. Artificial neural network was used for detection of breast cancer by Menéndez et al. [30]. They proposed a system for breast cancer diagnosis to determine which women are more likely to suffer from a particular kind of tumor before they undergo a mammography. Several SOM networks of increasing dimensions were trained with original set of 49,401 healthy patients. Afterwards, an algorithm of unsupervised learning and low computation cost (MARS) was used in order to train a different model for every SOM network trained. Testing of trained network was done again on three different datasets which yielded promising results.

Input: - The value of pixel p_t in frame where t=0….lastframe
Output: - Binary mask value $B(p_t)$ of foreground/background
Variables: C(current pixel model), A(array), p(pixel), c_m (best matching weight vector)
Begin
Step 1 : Initialize the current pixel model C for pixel p_0 and store it into A.
Step 2: for t= 1 to lastframe
    Evaluate best matching weight vector c_m in C for current sample p_t
    if (c_m found) then
        $B(p_t)$=false;
    Update A in the neighborhood of c_m
    else if (p_t shadow) then
        $B(p_t)$=False;
    else
        $B(p_t)$=True;
    End if
  End for
End

C1: Pseudo Code for utilization of Self Organizing Map for Background update [31,33]

A SOM based algorithm for modeling difference between stopped and moving objects captured by stationary camera was proposed by Maddalena et al. [31]. SOM was used to develop a model of scene background as well as foreground that can handle critical situations such as dynamic backgrounds or rapid illumination changes etc. The technique was employed to help differentiating moving and stationary foreground regions, which further leads to initial step for segmentation of objects in a scene. The scheme was tested on superimposed objects also. Another SOM based dynamic retinotopic method for enhancing capability of moving object detection in videos was proposed by Quintana et al. [32]. Retinotopic SOM was a visual vertex mechanism that allowed its learning parameter to adapt based on scene behavior and mimics perception abilities. Standard SOM architecture was used and involves Hebbian learning mechanism for updating neuron weights. The method could deal with dynamic background and showed robustness with illumination changes. Furthermore, the method could predefine parameters automatically which made it suitable to use in real time applications for video sequence analysis. Maddalena et al. [33] proposed another Self Organizing Background Subtraction approach for tackling issues such as sensitivity to dynamic environment, changes due to lightening and extraneous events present in traditional Background Subtraction method. In their work, background motion trajectories were learned in self organizing manner to help in detection of motion, updating the background model and further, making neural network architecture much simpler. They organized ANN as 2D flat grid of neurons where weights of the input nodes represented neural network learning (Please refer to C1 for pseudo code and detailed algorithm). The proposed method acted as competitive neural network, as each node was mapped to most similar incoming pattern. But in case of color pixels, a neuron map of n*n vectors was taken and mapping of incoming pattern was done using the nearest distance measure to the weight vector and further, neighborhood weight vectors were also updated. The whole set of weight vectors behave as a background model that is used for background subtraction. The performance of proposed approach was evaluated using frame based and pixel based metrics and results showed accuracy with lower computational cost. A different background subtraction approach based on Self Organizing Neural Network architecture known as temporal modular self-adaptive SOM was proposed by Alonso et al. [34]. The basis of approach was that it could adapt to two different Background Subtraction models and incorporate an automatic behavior in the system. SOM was used to decide which Background Subtraction model shall be used based on dynamic system behavior for improved segmentation results. Neurons were updated using input frame information and finally average filters were used to classify pixels as foreground or background by considering neighborhood information. Four specialized modules were included to handle various scenarios that can occur in videos and make decisions accordingly. The work was evaluated on some video sequences of Change Detection [179], BMC [178], and Wallflower database [166]. A method based on 3D self organizing background subtraction (3dSOBS) algorithm was elaborated in another study by Maddelena et al. [35] to generate background model automatically for detection of moving objects. This algorithm enhanced previous self organizing background subtraction algorithm by incorporating initial background estimation, spatial coherence, and shadow detection and removal methods. In their work, network was organized as 3D grid of neurons, the background model was initialized by the first frame of the video and occlusion was handled using temporal median method in initial frames. The weights were updated using a visual attention mechanism of reinforcement and exploiting spatial coherence of scene objects. A threshold calculated empirically was used to classify pixels as foreground and background. They extended the method adopted by [33] for discriminating shadow pixels by taking into consideration the fact that shadow cast on background does not have any significant change on hue. Experimental results performed on extensive Background Models Challenge datasets demonstrated that 3dSOBS algorithm handles situations such as waving trees, cast shadow for tiny and large moving objects robustly than other state-of–the-art methods. Ikeda et al. [36] proposed another ANN based system for counting pedestrians in single image captured by single video camera. Background Subtraction algorithm was applied to extract group of pedestrian between input and background image. The color and shape information was calculated for every pedestrian region that constituted a nine dimensional vector. Back propagation algorithm was applied to estimate number of pedestrian in group and tracking was accomplished by comparing current location of detected pedestrian with previously extracted location.

Convolution neural network is one of the most notable deep learning [37,38] approaches employed for various computer vision tasks such as object detection, image classification and object segmentation. For pedestrian detection, Tome et al. [39] have proposed architecture based on deep learning that adapts convolution neural network for tasks at hand. The proposed approach consisted of three stages: region proposal, feature extraction and region classification. They had tested three different strategies (Sliding window, Selective Search and Locally Decorrelated Channel Features) for region proposal. In the context of object detection, convolution neural network was trained on ImageNet dataset. Region proposal score was used to select regions which were fed to network for classification. It also exploited an additional feature which was used as final classifier for detecting the presence of a pedestrian. Such image classification approach outperforms state-of-the art in terms of detection accuracy based on handcrafted features and low computational cost. Due to high computational complexity of deep neural network, there is a strong demand for reliable and efficient object tracking algorithm. To address this problem, Jiang et al. [40] came up with convolutional based pedestrian detector by sharing features across a group of deep neural Network based detectors having various sizes of model windows. They detected pedestrians of various scales on one single layer of an image pyramid simultaneously. The proposed approach was able to decrease number of layers that were required to take out features compared to traditional scanning approach. They used four convolution neural network detectors to make detection performance stable for further reduction of computational burden. For object detection system, object classification in videos has been investigated intensively due to its important role in various applications. To improve object classification, Najva et al. [41] introduced a model for detection and classification of objects in videos based on deep neural network. In this approach, object detection was done using background subtraction approach. Object recognition and classification was done using feature vectors including tensor based Feature Extraction, along with scale invariant feature transform to classify the detected objects using Deep Neural Network. Auto encoders were used to train deep network. Another relevant approach to infer impact of object motion in videos was proposed by Rodriguez et al. [42]. The proposed scenario included prediction mechanism to identify object motion in videos based on stored information in the neural network structure with characteristics of objects. The architecture has been implemented on GPUs due to parallel nature of neural network. The suggested method was evaluated on CAVIAR dataset. Braham et al. [43] proposed convolutional neural networks based background subtraction algorithm to enhance efficiency of object detection in videos. The method used a single gray scale image background model and convolutional neural network was trained using scene specific dataset. The aim was to design an algorithm capable to encompass deep feature classification potential and learn with convolutional neural networks to improve background subtraction operations. By testing on the ChangeDetection.net dataset (CDnet 2014), the authors showed that the detection accuracy of the proposed algorithm was better than state-of-the-art methods. One recent approach based on deep learning for brain tumour detection has been proposed by Havaei et al. [191]. A new Deep Neural Network architecture which exploited both local features as well as more global contextual features simultaneously has been proposed. The deep network implemented a final layer that was a convolutional implementation of a fully connected layer to aim speed up than traditional methods.

3.2. Neural network based object tracking

Neural networks have also been used for tracking of moving objects and this section summarizes available studies in that direction. A Growing Competitive Neural Network (GCNN) method derived from classical competitive neural network approach was presented by Lobito et al. [44] for tracking moving object for video surveillance. GCNN was introduced when creation of new process units (neurons) was a necessity for providing a better foreground object representation. It was observed that architecture was able to meet set of requirements such as time constraints management, robustness, high processing speed and re-configurability. A parallel and independent tracker known as MS-RBF(Mean-Shift Radial Basis Function neural network tracker) was proposed by Rowghani et al. [45]. The method combines two independent trackers where MS tracker performed target modeling using spatial-weighted color histogram and Bhattacharyya coefficient was applied for computing similarity measure between reference and candidate models. Radial Basis feed forward neural network tracker learned target features using supervised and unsupervised steps. It had input feature vector which learns color and texture features both from background and target, which was further used for detection and tracking of objects in other frames. A novel activation function was employed and the extracted features were clustered using Fuzzy C-Means clustering. The results of both trackers were averaged to get final output.

Fan et al. [46] introduced method for tracking some specific objects with a focus on human tracking. The proposed approach was based on convolutional neural networks in which spatial and temporal features were learned during offline training process. The use of conventional convolutional neural network was extended by applying shift-variant architecture to combine global and local features in a natural way. Initially, CNN extract local features from input image and were combined to obtain higher order features by further subsequent layers. CNN used shift-invariant property for detection and tracking was performed on the basis of spatial and temporal structures of the patches. The spatial structure was concerned with appearance and motion information was captured using temporal features. The output of tracker was probability map where location of target was indicated by a peak of the map. A robust online adaptive tracking system was developed with help of two neural classifiers (object/non-object) by Babu et al. [47]. Features such as color and location information of background pixels and surrounding objects were used for training network. The target of separating object from background was accomplished by estimating likelihood map among these pixels. The classification accuracy was increased further by tuning object and non object classifiers using fast learning radial basis function network for training of both classifiers. Background pixels were maximized using background classifier and classification of object pixels was maximized using object classifier. The shift in centroid of the object pixels specifies the displacement of the object and the cumulative displacement specifies the location of the object for a particular frame. To handle challenges like change in appearance, background and rapid illumination changes, the output weights of object/non-object classifier were adapted online. The intersection of posterior probability of two classifiers was used to derive both target and object model. Asvadi et al. [48] also used color features to evolve an object model based on fast learning radial basis function networks(RBFNN). The work focused on developing a background model for object of interest starting from initial frames and selecting objects to be tracked manually. K-means segmentation was then applied for separating the object from background and the background modeled in previous stage was extended. In the next stage, RBFNN was trained using both extended and object background features and object was tracked by iteratively seeking the object's centroid estimated by RBFNN method. The stopping condition occurs when centroid location remained unchanged for any two consecutive iterations. The proposed approach was evaluated using video sequences of PETS 2001 dataset.

An organized snapshot of the methodologies adopted to handle sudden illumination changes, dynamic changes in object was presented by Suresh et al. [49]. They applied online learning radial basis function classifier for object detection in videos. The hidden neurons were automatically determined by learning algorithm for capturing the decision boundary. To handle rapid illumination changes, parameters of the radial basis function network were adapted by online learning from neural classifier. The target objects were localized in the consecutive frames using posterior probability map. The challenges such as low quality of camera, Pan-Tilt-Zoom operations and various complex working environments possessed by real-world CCTV footage makes it hard to track objects accurately. Aiming to solve this problem, Dimou et al. [50] introduced a deep learning based multiple target detection framework to track objects in videos. The method incorporated heterogeneous training data and data augmentation for improving detection rate in CCTV scenes. Recurrent Neural Network was applied to model objects' spatial transformation parameters and to predict evolution of intrinsic camera parameters. The predicted parameters were helpful to tune detector parameter which further improved tracking results.

Tracking multiple people in videos is a challenging problem which is gaining attention especially for video surveillance applications. To tackle this problem, Xue et al. [51] came up with a deep learning model trained using big data acting as an efficient classifier for different types of object. The focus of study was on pedestrians tracking in RGBD videos captured by RGBD camera. Deep convolutional neural network classifier was trained offline and procedure for tracking was run online to further boost performance for human classification. A probabilistic tracking method was introduced to integrate deep learning based classifier with a model of probabilistic motion in which people were tracked by increasing posterior probability greedily. Due to powerful feature learning capabilities, deep learning has been successfully used for visual target tracking but trackers suffer from contaminated features obtained in occluded areas when occlusion occurs. Therefore for tracking moving target in occlusions, Wu et al. [52] proposed regional deep learning tracker in which targets were observed in multiple sub-regions. The stability of these sub-models was characterized by devising a stable factor which was modeled as a hidden variable of Factorial Hidden Markov Model. The factor was able to give a confidence degree during inference stage for response score of every model and also helped in evaluating online training criteria for every deep learning model. This training (online) approach also enabled acquiring of various local features accurately as compared to fixed training trackers. The weighted Gaussian Mixture Model under particle filter framework was used to approximate tracking results to improve computational efficiency. Wang et al.

[194] proposed visual object tracking method using deep learning technique. The proposed tracking method comprises of offline training stage and online tracking stages. They trained a stacked denoising auto-encoder(SDAE) offline using unsupervised learning for generic features which was further transferred to online tracking task. Tiny Images dataset was used for offline training. Layer to layer training was applied initially and then SDAE was finely tuned. A sigmoid classification layer was added to the encoder part of the trained SDAE during online tracking process resulting in classification neural network. The object to be tracked was represented by bounding box in the first frame. The particles were drawn according to the particle filter when the first frame arrives. The confidence of each particle was activated by making simple forward pass through the proposed network architecture. Whenever maximum confidence of all particles in a frame falls below a threshold, it indicates significant change in the appearance of the tracked object which was handled by tuning the whole network again. Researchers had often used object appearance models for visually tracking object in videos. Chen et al. [53] proposed a deep learning framework to learn discriminative features via a convolutional neural network (CNN) to improve tracking performance. They designed an object appearance model using hierarchical feature extraction approach and a likelihood function was used to estimate appearance variation. Finally, a heuristics schema was applied to evaluate whether updating of the object appearance model is required or not. Experiments were done on CVPR2013 dataset to prove tracking efficiency of the proposed method. (Please refer to C2 for pseudo code and detailed algorithm).

Initialization:
(i) Deep convolutional neural network was pre-trained on CIFAR-10 dataset.
(ii) Acquire manually labels for the first frame.
(iii) Positive and negative samples were collected and the corresponding image patches were cropped out.
(iv) Positive/negative image patches were resized to 32*32 pixels.
(v) The pre-trained Deep CNN-based appearance model was finely tuned.
(v) At time t=1, the particle set was initialized.
(vi) The likelihood threshold and maximum buffer size for set of online positive samples were initialized.
Tracking:
For iteration =2 to the end of the video
1. The candidate samples were generated based on previous particles.
2. Likelihood evaluation between candidate samples and target appearance model was computed.
3. The posterior density of target state was estimated.
4. The posterior density was reformulated using Bayesian theorem.
5. The object state was determined by the maximum posterior estimation.
(i) if the posterior density of the optimal object is within predefined threshold values then
    (a)At time t, positive and negative samples were collected respectively.
    (b) The set of online positive samples were updated.
    (c) If the size of set of online positive samples > predefined threshold then set of online positive samples was truncated.
    (d) The final positive sample set is updated.
    (e)The deep CNN-based appearance model was updated based on set of positive samples and set of negative samples.

C2: Pseudo Code and Methodology of CNNTracker [53]

Another technique which focused on visual tracking method based on deep Network in Network framework was proposed by Chen et al. [54]. The method learns discriminative features in a data-driven way automatically and do not rely on engineered features. A simple as well as an effective method for transferring generic and mid-level features learned from deep network for tracking task was designed. The drifting problem was addressed by utilizing samples collected in initial and the most previous frames. The proposed approach was experimented on CVPR2013 dataset to evaluate its tracking performance in videos. The method presented by Zhang et al. [154] applied deep learning driven approach for block-wise detection of moving objects based on binary scene modeling. The method developed deep feature representation by learning a stacked de-noising auto-encoder. Thereafter, the learned feature representation was mapped into the binary Hamming space. Finally, a density analysis based scoring function was developed for detection of moving objects. (Please refer to C3 for pseudo code and detailed algorithm).

The authors of [195] introduced visual tracking system based on pre-trained Convolutional Neural Network(CNN) for image representation and target localization. They initially applied pre-trained CNN for extracting features for given set of samples on the input frame and an online Support Vector Machine (SVM) was used to classify them. The relevant features were back projected to target for each positive sample. These were determined by examining SVM parameters through the network to generate saliency map for bringing out regions to distinguish targets from background. The target-specific saliency map was built by aggregating positive examples. Sequential Bayesian filtering was used to perform tracking using this target–specific saliency map. A generative model was shaped from target appearances accumulated in the previous saliency maps and convolution between appearance model and target-specific saliency map was computed by dense likelihood map. The SVM and generated models were updated based on tracking results of the current frame.

Input: Video sequence(X1, X2....)
Variables: parameters for feature extraction: Weight(W), Bias(b), threshold for detection(T), binarization multiplier(d), number of layers for feature extraction network(k).
Output: Foreground object mask(S_1, S_2.....).
Step1: initialization of scene model
For iteration =1,2,3....do
Step 2: Represent the present frame X_n with distinct blocks i.e. block at location l.
Step 3: Perform feature learning using Stacked Denoising Auto-encoder Learning
Step3.1 Input data was perturb with Gaussian noise
For iteration =1: k-1/2 do
Step3.2 Initializes weight (w) and bias (b) parameters wi, bi w(k-i), b(k-i)
Step3.3 Xin <---X` and Xtarget<--X;
Step3.4 Learn w^i, b^i $w^{(k-i)}$, $b^{(k-i)}$ by solving
Step3.4.1 $\min_{w,b} J(w^i, b^i w^{k-i}, b^{(k-i),} X_{target}, X)$ with $X=f(w^{(k-i)}Y+ b^{(k-i)})$ and $Y= f(w^iX_{in} + b^i)$
Step3.6 X_{in}<---Y, X_{target}<-- Y
Step3.7 end
Step4: Feature binarization based on hashing was done.
Step5: Background classification was done based on the estimated threshold (threshold estimated using mean and variance respectively.)
Step6 : Scene model update, if necessary.

C3: Pseudo code for Deep learning driven block-wise moving object detection with binary scene modeling [154]

Zhang et al. [196] introduced discriminative model using Convolutional neural network for object tracking in which image representation and localization of target was done using pre-trained CNN. Features were extracted using pre-trained CNN for given input frame of the video and online SVM(support vector machine) was trained to classify them. The features relevant to target identified by observing SVM model parameter were back-propagated for positive samples. These were used to obtain a saliency map highlighting regions discriminating target from background. The target saliency map was developed by aggregating positive example saliency map. Sequential Bayesian filtering was employed finally to perform tracking using target saliency map as observational model. They trained CNN with no extra dataset for performing object tracking. The tracker was made to adapt to various dynamic variation of background and object by employing several iterations of back-propagations. In [197], Zhang et al. combined convolutional neural network and random vector functional link network to improve and simplify the visual tracking performance of the tracker. The filters in the CNN were fixed and initialized randomly. The proposed model did not need any pre-training but it only required to train the parameters in the fully connected layers. Those parameters were updated using recursive least squares. There were 100 filters banks in the convolutional layer. The image patches cropped from positive and negative samples will act as input.

The convolutional layer used random filters for convolving the input. The features were normalized in a feature sliding window manner by 2D pooling layer. The fully connected layer which is the last layer will connect the input and output of the last normalization layer with the output neuron. Only translation was considered for CNN tracker to increase its efficiency. The target search was conducted on a Cartesian grid of unit step within a radius of previous prediction. The prediction position of the object signifies positive sample and the local image patches that do not significantly overlap the prediction resulted into negative sample. Wang et al. [198] presented some CNN properties with respect to visual tracking and also introduced full connected CNN based algorithm which was pre-trained on image classification task. They selected the relevant feature map by performing feature map selection process for given target on the conv4-3 and conv5-3 layers of the VGG network (Visual Geometry Group). A general network (GNet) was created on top of the selected feature maps of the conv5-3 layer which captures the target category information. A specific network (SNet) was built on the selected feature maps of the conv4-3 layer to perform distinction between the target object from background. In the first frame, both GNet and SNet were initialized for performing foreground heat map regression for the target. A region of interest (ROI) which centered at the last target location for a new frame was cropped and propagated through the fully convolutional network. This leads to generation of two heat maps which were further used to perform target localization. At last, disaster detection scheme was used to decide which heat map can be used to determine the final target. A Recurrent neural network(RNN) contains minimum one feedback connection which enables them to perform temporal processing and learn sequences. The authors of [199] claim that its work is the first one to introduce multiple target tracking based on recurrent neural network. The proposed approach was divided into two blocks: state prediction and update, as well as track management on one side and data association on the other. The dynamic model of targets was learned using temporal RNN. The RNN gives four values as output: vector for predicted states for all targets, vector defining all updated states, probabilities vector defining the existence of real trajectory and the absolute difference to the previous probability. The decision whether target exist depends on current state, existence probabilities, measurements as well as data associations in the subsequent frames. The target was captured using an extra variable whose values is between 0 and 1 to imitates the probability concept where 0 means target does not exist and 1 means target exists for a particular time instance. The testing of the approach was conducted on targets for threshold = 0.6.

3.3. Discussion

Table 2 presents a quick summarization and analysis of available neural network based studies on moving object detection and tracking. The table defines problem tackled by the study, reasons for employment of neural network based solution and architecture employed by study together with some important features of the paper. It can be observed that there is relatively good number of studies based on unsupervised learning as compared to supervised learning as studies on self Organized Maps (SOM) and competitive neural network are widely available. This is understandable as impreciseness and fuzziness of real time conditions is major hindrance in framing effective rule base for segmentation of foreground objects for supervised algorithms. In addition, system has to incur substantial overhead for training of the system and therefore unsupervised algorithms clearly have an advantage. Although the solutions offered by neural networks were appreciable but with advent of deep architecture and deep learning, focus

Table 2
Summary of neural network based methods for object detection and tracking in videos.

Problem Statement/	Paper	Why Neural Network?	Approach	Strong/weak points
Neural Network based detections				
Conditions like illumination changes, random noise affects detection process.	22	Neural network can be used for object classification because it is proven efficient tool for classification.	Feed-forward Network	Provides faster and better results than traditional BS as complex background generation wasn't required.
Movement and vibration of camera leads to noisy images making detection process complicated task.	23	The match and estimation task of Kalman filter can be assigned to MLP which can be trained using back-propagation algorithm.	Multi-layer Perceptron	The proposed approach made detections of object accurately than traditional approaches.
Sudden illumination and background changes affect learning of new scenes quickly and accurately.	24	Multimodal background could be represented by Unsupervised Competitive Neural Network(CNN).	Unsupervised CNN	The shadow detection method can be integrated in the proposed work to improve its performance.
Dynamic lighting, shadow and non static background lead to ineffective detection outcome	25	Integration of NN approach with Bayesian Classifier can help in reducing error and increase efficiency of system.	Feed-Forward Neural with Bayesian classifier	Proposed approach showed improvement in segmentations as compared to probabilistic approach.
Conventional Neural Network is unsuitable for real time applications due to long training time.	26	CNN is applied to perform parallelized object segmentation in complex real time video sequence.	Unsupervised CNN	The approach was parallelized on a pixel level and designed to enable efficient hardware implementation.
GMM was not able to capture intrinsic directionality of data at high computational cost.	27	Uses of dipoles in CNN effectively capture intrinsic directionality of data at low computational cost.	Dipolar CNN	Provided stable performance which makes it suitable for implementation in real world environment.
Standard detection method suffer from problems like spurious blobs, occlusion, annoying noise etc.	28	Self Organizing maps(SOM) is capable of handling such problems and detect anomalous object in videos.	Self Organizing Neural Network	The proposed work was computationally expensive.
It is very difficult to model complex environment	31	Visual attention mechanism can help in detecting human behavior features based on 3D self-organizing network.	3-D Self Organizing neural network	Stationary cameras resulted in gathering of imprecise and inaccurate information.
Self Organizing BS works fine with dynamic background but requires manual updation of parameters.	32	Dynamic Retinotopic Self Organizing neural network approach can be used to adapt its learning parameters based on scene behavior.	Dynamic Retinotopic SOM	Adjusting parameters automatically rather than predefining manually made it suitable for real time scenarios.
Critical situations such as illumination changes, random noise affects the detection process.	33	Self Organizing BS can deal with critical situations by generating background automatically without prior knowledge about involved pattern.	Self Organizing neural network approach	The proposed approach worked well with critical situations such as illumination changes.
Traditional approaches cannot handle severe illumination changes, stationary objects etc.	34	Temporal modular self-adaptive method allows system to handle various critical situations that leads to improve system performance.	Temporal modular self-adaptive SOM	Approach was able to handle problem such as drastic changes in environment and separately dealt with various segmentation issues
Insufficiency of SOM based BS fail to accurately detect object in dynamic environment	35	NN offers advantages such as parallelism, learning, adaptivity that can be exploited for smoother classification.	3dSOBS+ algorithm	The proposed approach was able to achieve very accurate definition of dynamic object.
Template matching scheme for Pedestrian detection in crowded group gives imprecise & inaccurate results	36	BS algorithm with Back propagation algorithm is able to detect pedestrians from crowded group of people accurately.	Back propagation algorithm	The proposed approach effectively counted pedestrian from video sequence images and provides flexibility to support complex actions.
Insufficiency of traditional pedestrian detection system builds uncertainty in classification task	39, 40	Deep learning based Convolutional Network can improve overall performance by optimizing most stages of traditional pedestrian detection pipeline	Convolutional Neural network	The results of classification modeling decreased by applying convolutional neural network approach.
Object classification is challenging due to factors such as object size, occlusion, scaling, lightening etc.	41	The system is able to classify object in video accurately by incorporating Tensor features along with SIFT using Deep Neural Network.	Deep Neural Networks	The system can be proficiently used for classifying object in videos.
Fast and accurate object motion detection pose a big challenge for applications such as surveillance, vehicle navigation etc.	42	A system based on Growing Competitive Neural network capable of representing motion under time constraints.	Growing Competitive Neural network	The system was able to meet a set of requirements such as time constraints management, robustness, high processing speed and re-configurability.
BS based on low level or hand-crafted features leads to ineffective object detection outcomes.	43	BS algorithm based on spatial features learned with convolutional neural networks was proposed to increase system efficiency.	Deep BS with Convolutional neural network	The proposed approach achieved promising detection accuracy and detected objects when scene specific knowledge was considered.
Neural Network Tracking				
Classical competitive neural network approach cannot manage variable number of objects	44	Growing competitive neural network can model uncertainties by generating new process units (neurons) when needed to get a better representation.	Growing competitive neural network	With a reliable tracking algorithm, system was able to determine whether there are suspicious/dangerous objects or not.
Difficulty in modeling dynamic trajectories by mean shift trackers.	45	Able to model dynamic trajectories and improves tracking performance By integrating mean shift trackers with radial basis neural network.	Mean-Shift Radial Basis Function NN tracker	The proposed tracker was able to track more than two objects simultaneously.
Deviation of objects from original positions led to failure of traditional approaches in real time situation.	46	CNNs learn both spatial and temporal features jointly two adjacent frames to overcome deviations.	Convolutional Neural Networks	Handling occlusions of objects in frames was real challenge and limitation.
Learning process of traditional NN is intensive and requires several hours.	47, 48	Robust object tracking algorithm using fast learning RBF networks gives a wider scope for modeling objects using neural networks.	Radial Basis Function network.	The proposed tracker is suitable for real time applications due to its low computational cost.

Table 2 (*Continued*)

Problem Statement/	Paper	Why Neural Network?	Approach	Strong/weak points
Conventional tracking algorithms fail under rapid illumination variation, appearance changes.	49	Online learning neural tracker (OLNT) adapts parameters of radial basis function network to handle changes in object/background dynamics.	Radial Basis Function network.	Required smaller computational effort but major limitation was its inefficacy to track fast moving objects.
PTZ operation, low camera quality, moving background motion blur, affect accurate multiple object detection in CCTV footage.	50	False detection pertaining to critical situation is eliminated by exploring heterogeneous training data and data augmentation using convolutional neural network.	Convolutional Neural network	The proposed approach was found to be effective and robust for both original and blurred content, as well as challenging action scenes captured in CCTV videos.
Use of Traditional classifiers and features can distinguish between object categories but faces difficulty in tracking task	51	Deep learning is able to extract representative features for images by training a pedestrian classifier for classification of persons in videos.	Convolutional Neural network	The method was able to achieve better performance in classification and recognition than support vector machine.
Various unreliable situations such as occlusion make it difficult to model trajectories.	52	Deep learning approach can be applied to achieve better tracking performance by integrating multiple sub-region observation models.	Regional deep learning	Deep learning based trackers gave better and stable performance by using hidden variable of Factorial Hidden Markov Model.
Definition of hand-crafted features requires expert knowledge and time consuming manual adjustments.	53	Deep convolutional neural network can learn discriminative features automatically.	Deep convolutional neural network	Online training with a multi-layer deep modeling is more powerful than tracking using handcrafted features and shallow models.
Conventional tracking method isn't able to handle object appearance variations in modeling complex dynamic trajectories.	54	Visual tracking method via deep Network in Network (NIN) can solve object appearance variation problem by learning most discriminative features automatically in a data-driven way.	Deep Network in Network (NIN)	The performance was better than as it could develop robust object appearance models by effectively learning and transferring highly general NIN features.

of researchers is eventually shifted towards exploration of such approaches to tackle challenges posed by detection and tracking environments. It is clear from number and results of studies that success of such studies had increased considerably. This is because of the fact that deep network based systems are able to successfully adapt to challenges like motion blurring, scale changes etc and can provide efficient classification results also [39–41]. It can be concluded by inspecting results (attribute based analysis) of deep learning based approach that performance of deep learning is quiet appreciable for following challenges/attributes: illumination variation, deformation, out-of-plane rotation, background clutter, occlusions and scale variation. The performance of deep learning is favorable for attributes like motion blur, fast motion, in-plane rotation and out-of-view but its performance is really demanding while handling low resolution attributes. According to precision and success plot (quantitative analysis) of related studies, deep learning approaches achieves overall good performance than state-of-the-art methods as observed in overall experiments reported for tracking and detection in videos. The results indicate that success of deep learning based method is still a challenge in low quality or noisy images and comparable results can also be obtained with other tracking mechanisms [52,53]. It is observed that majority of object detection and tracking studies in videos apply traditional convolutional neural network which learn spatial features from single images. Unfortunately, this feature makes it difficult for deep learning to achieve remarkable performance in this domain using CNN. It is envisaged to apply 3D CNN [55] in the future to carry findings because it has the capability to learn features of videos directly. In the future, researchers can explore other deep learning methods such as Convolutional restricted Boltzmann machine learning [190] for improving performance of object tracking in videos. Recent attempts like use of RNN [198] are also expected to be future candidate for triggering research in this direction. To summarize, deep learning is one of the novel frontiers which has the capability to improve performance of object detection and tracking in videos and a number of studies in this direction are expected in coming years. There is room for further improvement despite the presence of promising results in this domain. Further, research investigation for deep learning can be to explore field of medicine to develop object detection and tracking algorithms using suitable architecture. Improvement in computational complexity for deep learning networks is another area that can be targeted. There is need to focus on development of algorithm for real time applications and learning better feature representation approaches using deep learning. Another possible future direction can be to include implementation and testing of various other soft computing techniques in deep learning model and determine most efficient one to perform detection or tracking task. Combination of deep learning with evolutionary techniques to overcome challenges like over-fitting, dropout to avoid co-adaptability, unsupervised pre-training for weight initialization can also be exploited.

4. Fuzzy logic based approaches

Fuzzy sets can work in situations with imprecise information as is the case of detection of object in real life situations. Fuzzy sets have the capability to simulate human reasoning for modeling imprecise concepts. Researchers choose fuzzy logic because it is well calibrated with information which cannot be easily quantified or modeled. Since there is always fuzzinees involved with computer vision because of cameras or real life environments therefore writing precise mathematical models for vagueness in real life environment is not practical. Fuzzy logic efficacies have been largely exploited by researchers over years to propose solutions for detection and tracking of objects in videos and few of them have been summarized here. The basic structure followed for their organization is shown in Fig. 6. Object detection and object tracking based studies have been grouped separately for better understanding.

4.1. Fuzzy logic based detection approaches

Authors of [56], Kim et al. had utilized clustering-based feature, called fuzzy color histogram for background subtraction in dynamic texture scenes. In their work, background subtraction was done by calculating similarity between observed and model Fuzzy Color Histogram features, renewed by online update procedures. Baf et al. [57] made use of Type-2 Fuzzy Mixture of Gaussians Model (T2-FMOG) which was first proposed by Zeng et al. [58] to handle higher uncertainty levels/parameters present in multimodal environment in standard Gaussians Mixture Model(GMM) method.

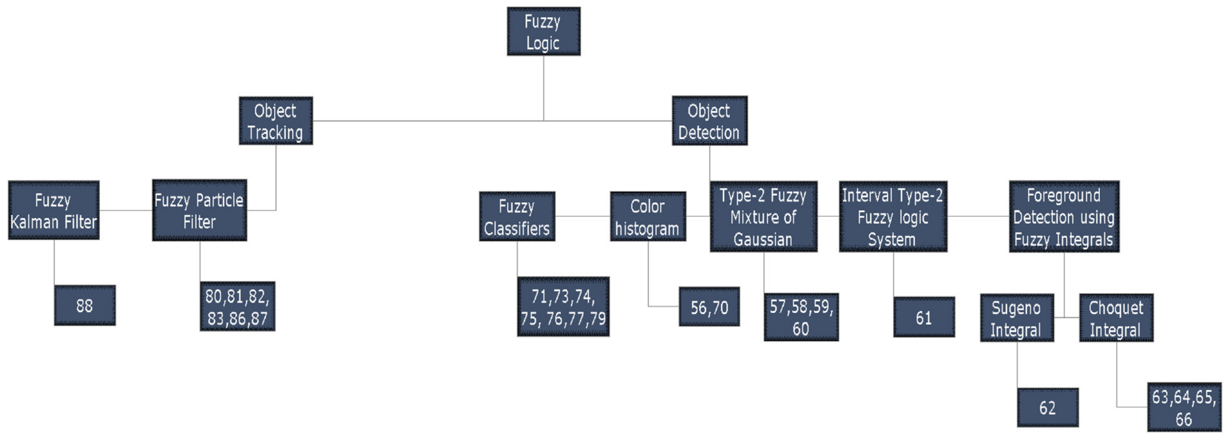

Fig. 6. Summary of fuzzy logic based object detection and tracking approaches.

Authors had presented two algorithms: T2-FMOG-UM to handle uncertainty over mean and T2-FMOG-UV for handling uncertainty over variance and had tested methods over indoor and outdoor videos. Bouwans et al. [59,60] had utilized type-2 Fuzzy GMM to represent background in order to improve performance and accuracy of standard background subtraction model. In [59] the authors had considered color features for detection of foreground object. They had taken three component color features from chosen color spaces of Ohta, HSV, YCrCb etc. A lambda fuzzy measure was used to compute fuzzy measures of all subsets of criteria. The foreground detection was made independently in each dimension and the color similarity measure was aggregated using Fuzzy Choquet integeral calculated in different dimensions. The choque integeral for each pixel was computed by rearranging the pixel in each set. The pixel was considered as foreground if its choquet integeral value is less than a certain threshold. Fuzzy adaptive background maintenance for modeling background was applied in which Choquet integeral results were used in the background maintenance update rule. In [60], type-2 FGMM-UM (uncertain mean vector) and type-2 FGMM-UV(uncertain variance vector) for modeling background of RGB videos was used. The means were only estimated as variance and weights remains unstable during GMM maintenance. During training of Type-2 FGMM, GMM parameters were estimated using EM algorithm. The foreground detection was done by ordering of K-Gaussians which assumes that background pixels correspond to high weight with a weak variance. Therefore, the Gaussian distribution exceeding a threshold was considered for background distribution and others represented foreground distribution. Log-likelihood was used to perform matching with incoming frames of videos. The pixel was classified as background if it belongs to Gaussian distribution identified as background and otherwise it was classified as foreground. Yao et al. [61] devised interval type-2 fuzzy logic system for improved silhouette extraction in dynamic real world environment. The input and output of fuzzy logic system was represented using interval type-2 fuzzy sets. The source images were captured using stationary cameras and were analyzed using GMM for foreground detection. The detected foreground by GMM was then partitioned into n*n blocks and then Global nearest neighbor was used to perform human tracking. Moving objects from human silhouette was detached based on partitioned foreground and obtained human centroids.

Zhang et al. [62] applied color and texture features to calculate similarity measure between current and background pixels. These similarity measures were then aggregated by applying Sugeno integral while Baf et al. in [63] had used Choquet integral for aggregation of same features to handle uncertainty in classification or localization of pixels. A fuzzy adaptive maintenance rule was proposed to handle uncertainty of the classification. Two more foreground detection methods using Choquet integral were proposed in [64,65] where as foreground detection using Sugeano integral gained less attention as one of the difficulties related to this fact was that measurement scale assumed by Sugeano Integral was ordinal [66,67]. Balcilar et al. [68] has worked on detection of moving objects from traffic surveillance video frames. They had also used fuzzy Background Subtraction algorithm using choquet integral to detect and eliminate uncertainties present in conventional Background Subtraction algorithm. Low rank weightage and fuzzy aggregated multi-feature similarity method was applied by Gayathri et al. [69] to detect moving object in dynamic environment. They had employed both Sugeno and Choquet integral to detect fuzzy similarities from ordered similarity function. A Fuzzy Color Difference Histogram Based Background Subtraction approach was introduced by Panda et al. [70] for detection of moving objects. The method applied color difference histogram (CDH) in the background subtraction algorithm by computing color difference between a pixel and its neighbors in a small local neighborhood. They utilized Fuzzy c-means clustering and a c-dimensional histogram to propose a fuzzy color difference histogram. The background for the proposed approach was modeled using first frame of the video and was updated using features calculated from the current frame(Please refer to C4 for pseudo code and detailed algorithm). The similarity between the current frame and the FCDH of the background model was measured by histogram intersection which was binarized further to classify pixel choosing appropriate threshold. The most relevant contribution of color difference histogram was that it was able to reduce errors that occur due to non-stationary background, rapid illumination changes and camouflage. The results of approach were tested on various complex scenes of some benchmark publicly available video sequences.

Fuzzy component based object detector was proposed by Iqbal et al. [71] for object detection. The aim was to design fuzzy geometrical configuration classifier to select candidates for object components that satisfy constraints put on by geometrical model of object and have a reasonable appearance match with respective components. Multi-level Gaussian pyramids were applied on each test image to decompose them. Human-centric hybrid fuzzy classifier proposed in [72] was applied to each level in the Gaussian pyramid to generate candidate for all the components. Multiple candidates are retained and passed on to the fuzzy geometrical configuration classifier along with the appearance match scores. The appearance match scores were computed from individual component detectors for each candidate using fuzzy geometrical configuration classifier and further combined with the geometrical model of the object computed during the training phase.

The devised method was immune to occlusions. Application of fuzzy rule-based classification scheme on moving blob regions for human motion detection in videos was introduced by Balakrishnan et al. [73]. Fig. 7 depicts its complete methodology in which a background model was developed and classification features were extracted in RGB color space. Chromaticity and brightness distortions features were used for classification of background and foreground. Morphological operations and blob identification were two tasks of preprocessing stages along with use of median filter for removal of noise. Features of blobs were extracted and fed to rule based classification engine to classify extracted major blobs. Experiments were carried out on PETS2006 Database to evaluate efficiency of this approach. Another fuzzy rule-based technique used for detection and classification of changing regions in outdoor videos was adopted by Zeljkovic et al. [74]. The input patterns were extracted from the low-level description defined using MPEG-7 edge histogram descriptor. 80-bin histogram was used by this descriptor for representing the local distribution of directional and non-directional edges. They partitioned every video frame into 16 sub-images which were classified using fuzzy rule base system. They utilized multiple features to define a fuzzy function for approximation of every learning function with fixed rules for detecting target objects. The fuzzy classification system consists of three stages: Fuzzification, inference and defuzzification. During fuzzification stage, the feature values are transformed using membership function from continuous domain to discrete domain. The class for these features was detected during inference stage by combining fuzzy sets and fuzzy rules using T-norm operator. T-norm operator was used if a class appears in more than one rule. Defuzzification was finally applied to convert the results from fuzzy domain to real domain. Therefore, the final outcome was decided by adding the number of sub-images classified as building in single frame. They performed experiments using video test corpus of TRECVID forum to test their system. A set of geometrical extraction algorithms for developing behavior patterns was proposed by Arguedas et al. [75]. A behavioural fuzzy classifier was designed using type-1 and type-2 fuzzy logic to categorize behavior patterns into predefined semantic categories. The surveillance media management framework was proposed by authors to evaluate behavioral fuzzy classifiers against binary classifiers and behavior features against conventional appearance features. The algorithm was validated on AVSS 2007 surveillance dataset.

Input: let data X={x_1, x_2, ….x_n} where x= (i,j,k); i,j,k=0,1,….W-1, initial cluster centers V={v_1, v_2…., v_c}.
Output: Resultant Fuzzy membership matrix
Set iteration=1 to NumberOfframes do
Step 1: Calculate color difference histogram.
(i) Color images were quantized to W levels.
(ii) Convert RGB color space to CIE $L^*a^*b^*$ color space.
(iii) Local color difference histogram was calculated.
Step2: Compute Fuzzy color difference histogram
If (iteration==1) then
Step 3: Initialize background to
else
Step 4: Foreground detection
(i) The similarity function using normalized histogram intersection was calculated
(ii) Threshold the similarity function value.
Step 5: The background model was updated for background labeled pixels.
end if
end for

C4:-Pseudo Code for utilization of Object Detection using fuzzy color histogram [70]

A major challenge faced by the researchers is to detect and track objects in presence of rapid illumination changes. In this respect, Milanesi et al. [76] performed a classification based on fuzzy logic to identify moving regions that corresponds to unexpected objects in the scene after preliminary analysis and clustering of temporal changes in video sequence. They presented proposed approach using a three step strategy. Former, they performed change detection by taking into account the differential change between current frame and previous frame and absolute change between current frame and reference frame against threshold value. If this value exceeds a threshold then change point was labeled. This threshold value was computed based on the local average and standard deviation of the samples extracted with the mask. In the next step,

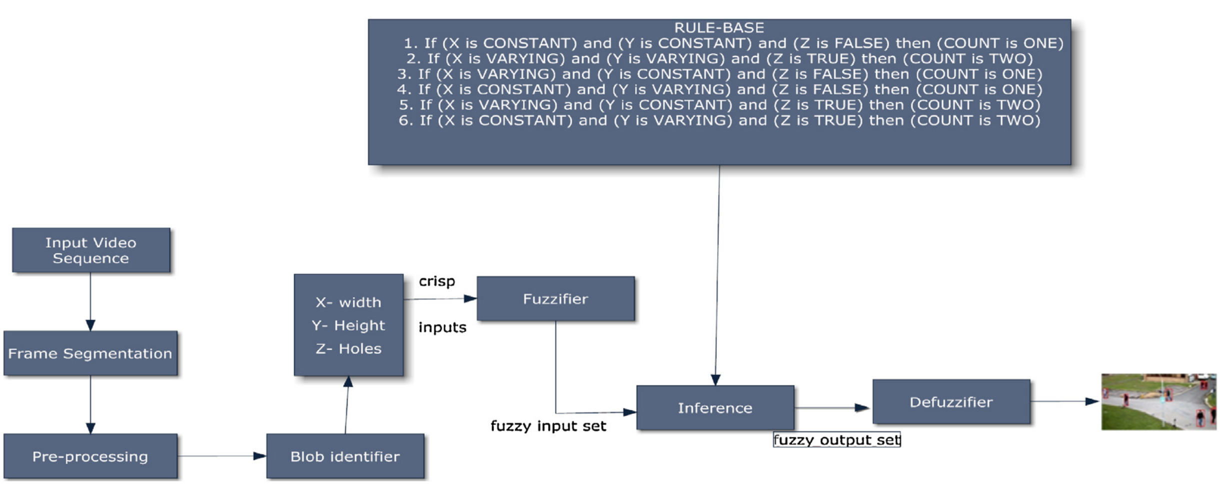

Fig. 7. Fuzzy inference System used for Fuzzy Rule-based Classification of Human Tracking and Segmentation [73].

tracking was performed based on three consecutive frame i.e. previous, current, and next, further correspondence between bounding boxes was determined using minimization of an appropriate cost function. At last, the tracked changed regions were distinguished from geometrical regions and from other type of regions using two parameters i.e. weight ratio and morphological index by defining semantic rules based on fuzzy logic. The membership function was determined by statistical analysis of the input linguistic variable. The proposed approach was evaluated on grayscale videos captured by a low quality webcam and a digital camera of good quality. Event detection and annotation of soccer game video on broadcast was done by Hosseini et al. [77] with help of fuzzy rules. This method presented a fuzzy rule-based classifier which applied statistical information obtained from set of audiovisual features and produced semantic concepts corresponding to the occurrence of events. The training data was used to derive continuous feature vectors based on discretization and fuzzification information. A decision tree was constructed to extract hidden knowledge among tuples and correlation between features and related events. A traversal of decision tree generated set of required fuzzy rules which were employed by fuzzy inference engine for decision making. Tests were completed on broadcast soccer game videos to demonstrate performance of the proposed method. A hybrid approach for human action detection was introduced by Vats et al. [78] that combined computer vision as well as fuzzy set theory benefits. The proposed work was based on Bandler and Kohout's sub-triangle product. The object window for given input video sequence was represented as a covariance matrix of features. The hybridization was performed on the tracking output in which covariance tracking algorithm was modified to perform part-based human motion tracking. The fuzzy qualitative translation states were used to represent the rigid motion of each body part. The semantic relationship between human and action was defined in terms of early detection of human actions. A frame-by-frame membership function for each kind of possible movements was constructed and detection of target was triggered when a pre-defined threshold was attained in a suitable way. They experimented on videos from Weizmann human action dataset. Ghasemi et al. [79] proposed a novel combination of fuzzy inference system and Dempster–Shafer Theory for the purpose of segmentation of brain MRI images where the pixel intensity and the spatial information are used as features. In the proposed modeling, the consequent part of rules is a Dempster–Shafer belief structure. Fuzzy Naive Bayes method were employed for calculating the basic probability assignment of focal element and dempster's rule was used for aggregation of FIS.

4.2. Fuzzy logic based object tracking approaches

Object tracking is one of the most critical tasks in video analysis applications. Two types of filters namely particle and Kalman are very commonly used to track object over frames in videos. Kalman filters are linear quadratic estimator and have low computational requirements. The particle filter provides a better way to track object as compared to Kalman filter. The Particle filter can be employed if system is not linear and changes with time. Kalman filter predicts current state from previous using system model and sensor observations. Particle filter generates different states randomly and assigns weights to those states according to evidence. States with fewer weights are discarded and new states are again added randomly. The Particle filter shows more promising results than Kalman filter because it provides supports for multiple predications to recover from lost tracks which make way for accurate tracking of complex object motion. However particle filter has a tradeoff between computational cost and system precision i.e. increase in number of particles also leads to increase in computational cost and vice versa. In order to handle these problems fuzzy based Particle filter [80] was proposed. The method performed modeling of uncertain variables with imprecise knowledge using fuzzy set theory. A number of fuzzy based particle filer trackers have been proposed in recent past and few are summarized in this section.

The authors in [81] proposed a fuzzy set theoretic particle filter approach for tracking multiple objects in varying illumination conditions for video surveillance application. They used adaptive Gaussian mixture model for modeling pixel at a particular location of the video frame. The probability distribution was estimated using weighted samples in particle filter. The randomly generated particles with uniform sets were used as initial set. Fuzzy based color histogram was used to establish one to one correspondence between frames for tracking in multiple object scenarios. After the object detection, its corresponding features such as area, centroid, etc. were calculated. The first location detected as foreground object was considered to initialize the search in following video sequence. The variations in the object detected area were captured using rate of change of elliptical major and minor axis. A fuzzy similarity measure was used to perform measurement in feature space. They initiated tracks based on the computed object area. The states for multiple moving objects were updated based on fuzzy color histogram and fuzzy distance measure. In order to tackle system precision, various fuzzy techniques were constructed for color feature association. Finally, computation cost was also taken care of because color space and fuzzy c-means mapping were calculated prior to tracking. An adaptive fuzzy particle filter method was developed by Varcheie et al. [82] to track objects captured by PTZ camera. They modeled whole process as a servo control system which consists of five main blocks: initialization, object capturing, target tracking, position prediction and camera control. Optical flow was incorporated to detect moving objects. Initially targets are extracted at the initial pan, tilt angles and zoom, which was further considered as input to the fuzzy based target tracker. Fuzzy classifiers were used to determine output defined in terms of pan and tilt angles of the camera. Fuzzy membership functions were used to weight particle filter samples and were later applied to geometric and appearance features. Object modeling and tracking was achieved based on sampling around a predicted position obtained by a position predictor. This approach is used for video surveillance applications such as video conferencing and tracking a speaker during presentation. However it worked well only for smaller objects and can handle short term occlusions based on assumption that object stays in changing field of view.

Fuzzy Particle filter integrated with classical Monte Carlo algorithm was proposed by Lin et al. [83]. This approach could estimate a posteriori distribution of the target state to enhance robustness of tracking in rapid illumination changes conditions. Fuzzy based video sequence motion tracking has been proposed by Garcia et al. in [84]. The method composed of three phases: segmentation, tracking and analysis which were based on fuzzy logic technique to facilitate working with permanency value uncertainty due to image noise. During segmentation phase, fuzzy bi-dimensional rectangular regions were used to evaluate identical features between targets along image sequence. In tracking phase, association between different fuzzy regions was performed in every image over time. Lastly, analysis phase was used to describe motion throughout a video sequence. During segmentation and tracking phases, possible errors that appear were corrected by fuzzy logic which provided a means of handling uncertain data. An optical flow based tracking algorithm and stereo vision in conjunction with fuzzy control to track object in the 3-dimensional space was proposed by Parrilla et al. [85]. This scheme has used optical flow algorithm for performing 2D tracking of the targets in a video sequence. The adaptive filters were employed to predict velocity values. Fuzzy controller was applied to amalgamate the velocities computed by optical flow algorithms and adaptive filter was used to predict object movement and approximate its trajectory when object disappears in video sequence. The input to the fuzzy controller was the absolute value of the difference between the velocity calculated by optical flow and velocity predicted by the adaptive filter. The other input was binary variable indicating the velocities calculated and estimated in previous frame. Fuzzy rules were framed by setting output variable high only for selecting the velocity calculated by optical flow when there is no occlusion and target is in the output of the obstacle otherwise output variable will be low for selecting the velocity calculated by adaptive filter. They employed center of sum method for calculating the output. Finally, when the targets were tracked in both the sequences then the 3D velocity of the object was calculated by triangulation principle. Another work using fuzzy particle filter was presented by Yoon et al. [86] in which adaptive model was implemented to improve performance of object tracking algorithm. The adaptive model comprises of two models: Adaptive AR model used as a state transition model and appearance mixture model being used as observation model. These adaptive models were not able to make optimal parameters when the state of a tracked object changed suddenly. To overcome this problem, fuzzy particle filter with fuzzy noises was applied to manage uncertainty in various noise models. Experiments were carried out on video sequences captured by camera in a vehicle under different environmental conditions. Liang-qun et al. [219] proposed visual multiple object tracking based on fuzzy logic approach. They incorporated fuzzy inference system for data association in multiple object tracking designed using fuzzy if-then rules. Rules were proposed for determining the fuzzy membership degree that substituted association probabilities between object and detection response (Please refer to C5 for pseudo code and detailed algorithm). The track-to-track association based on fuzzy synthetic was proposed to deal with fragmented trajectories. The efficiency of the proposed approach was evaluated on public datasets and also showed its ability to minimize the number of fragmented tracks.

A similar investigation was carried out on fuzzy particle filter for detection and tracking of various people based on color, stereo vision and fuzzy logic by Paul et al. [87]. They employed two fuzzy systems to filter out false positives of a face detector. The scheme evaluated whether the detected face has high possibility of the person to be tracked. In the next step position of person was calculated and occlusion map was updated. In case of non detection of any face fuzzy logic based particle filter was initiated which generated a sample set around the position of the object that was found and measured and weighted each sample based on this observations. New states were calculated further and weights were updated along with updating of occlusion map. This resulted into better tracking performance even when color information or stereo information was confusing in the reference camera image.

A fuzzy-controlled Kalman filtering scheme was proposed by Fernandeza et al. [88] for stereo visual tracking in videos. In this method, authors had designed two levels, firstly, a fuzzy methodology was introduced to estimate the plant noise covariance matrix over time instant to allow alteration of the filter to actual conditions; and secondly, fuzzy control stage based on a fuzzy feedback system was applied in those cases where a lost track was used to reinitiate the filter. The fuzzy inference was employed in a stereo-vision tracking system to adapt the gain of a Kalman filter by changing the process noise for actual conditions.

Step1: Initialization: To compute the detection response at the current frame the offline trained detector based on aggregated channel features[n2] was applied.
Step2: State estimation
For iteration =1,2….
Step2.1 Data association: Fuzzy logic data association algorithm was applied to determine the best association between measurements and objects.
Step2.1.1 Based on the affinities of motion, shape, RGB color histogram and local binary pattern, fuzzy inference system was defined to compute association between measurements and objects.
Step2.1.2 Two input variables $E_{i,j}(k)$ and $\Delta E_{i,j}(k)$ were defined in terms of the predication errors and change of errors of shape, motion and appearance models.
Step2.1.3 The universe of discourse and the range of values which $E_{i,j}(k)$ and $\Delta E_{i,j}(k)$ each may take is in the interval [0,1] regardless of the object types, Object-to-object occlusions and abrupt object motion.
Step2.1.4 five fuzzy sets that are labeled in the linguistic terms of zero (ZE), small positive (SP), medium positive(MP),large positive (LP) and very large positive (VP), were specified for each crisp input $E_{i,j}(k)$ and $\Delta E_{i,j}(k)$ in the fuzzy inference system.
Step2.1.5 The triangular and Gaussian membership functions was employed for representation of fuzzy sets.
Step2.1.6 The fuzzy rules can be expressed based on the input and output variables. Linguistic terms for fuzzy sets ZE(zero), SP(Small Positive), MP(Medium Positive), large Positive(LP), VP(Very Large Positive)
Rule 1: IF Ei,j(k) is ZE AND ΔEi,j(k) is ZE THEN ( ), μi,j(k) is EP
Rule 2: IF Ei,j(k) is ZE AND ΔEi,j(k) is SP THEN ( ), μi,j(k) is VP
Rule 3: IF Ei,j(k) is ZE AND ΔEi,j(k) is MP THEN ( ), μi,j(k) is LP ...
Step2.1.7 The Max-Min composition rule of inference technique was employed to calculate the crisp output and Center of Gravity(COG) was employed for defuzzification.
Step 2.2 States Updation: Kalman filter was applied to update the object states with associated measurements.
Step 2.3 Track Management:
Step 2.3.1 Fuzzy track-to-track association was used to associate the terminated tracks with new tracks.
Step 2.3.2 Track initialization and termination rules:
2.3.2.1 If the measurement is not associated with any existing tracks or new tracks, this measurement will be considered as a new track and initialized by using the Kalman filtering.
2.3.2.2When the new track is consistently associated in T_init frames, the new track is marked as an existing track.
2.3.2.3 If the existing track is not associated with any measurement in subsequent T_term frames, it is considered as the terminated track.
2.3.2.4 If the terminated track is not associated with any measurement more than λ frames, it is considered as the complete (ended) track.
End For

C5:- Online multi-target tracking based on fuzzy logic data association [219]

Optical flow [89,90] is one of the popular moving objects tracking algorithm for videos. But major challenge for researchers was removal of incoherent optical flow field. In this respect, Bhattacharyya et al. [91] proposed fuzzy hostility index to segment optical field for high speed target tracking in real time videos. In this method, the image frames are extracted from video sequence which was stored as bitmaps for optical flow computations. Then the optical flow between initial two frames was calculated to show variation in intensity structures between the frames. In the next step, flow of region of interest was determined by employing pixel hostility index. A thresholded optical flow field at a hostility index=0.5 was used. Regions were regarded as incoherent having fuzzy hostility index value greater than 0.5 and were filtered out from the flow field. A maximum density optical flow region was detected to find point of interest on the moving object to be tracked. The neighborhood of a predefined size for the point constitutes the region of interest (ROI) for operation in the subsequent frames. Therefore, tracking of the moving object will be shifting the region of interest with respect to optical field calculated around the detected point of interest in the subsequent frames. With many results from three sequences of video of high speed flying fighter aircrafts, it was demonstrated that method was effective for high speed target tracking in real time. The computation budget of optical flow field was reduced significantly. Three-dimensional feature and a fuzzy membership distribution composed of membership values was proposed by Kim et al. [92] to boost tracking performance of moving objects in videos. The proposed method consists of three parts. The first part was gate control to provide the location of the target by employing the predicted position of the tracking filter. The gate window was set as validated region for target extraction. The second part was concerned with preprocessing focusing on feature extraction from targets and distinguishing the targets from background noises. The features extracted during preprocessing were used for motion information estimation for the desired targets. Centroid estimation algorithm was used to determine the relative motion of the object using segmented target. The tracking algorithm outturned with tracking information such as predicted position and velocity of the target which was further put into to the gate control part for validated region setting. The authors presented state of the art results on various infra-red image sequences and CCD image sequences.

4.3. Discussion

Table 3 enlists various studies that had employed fuzzy logic principles for detection and tracking of moving objects in videos. The table specifies problem being targeted by these studies in first column. The table also lists reasons of why and how fuzzy logic was able to propose a solution for specific problem. The architecture employed by the study is also listed. Strong or weak points of study are also portrayed that can help in guiding future research. Fuzzy based approaches such as Type-2 fuzzy Mixture of Gaussian, fuzzy color histogram, fuzzy classifiers, fuzzy particle filters, fuzzy Kalman filter etc. has been effectively applied in numerous ways for object detection and tracking in videos. One can notice from studies listed in survey that fuzzy approaches are successful in handling uncertainty that is present in each and every level of object detection and tracking process in videos. The primary factor that inspires researchers to apply fuzzy approach in the domain of object detection and tracking in videos is their capability and flexibility to adapt to different types of uncertainties. The poor performance of conventional mathematical operators such as AND, OR, etc used for aggregation of rules for making pixel classification decisions has led to demand for utilization of fuzzy integrals [62–65] to improve crisp decision problems. Gaussian mixture model is one of the popular approaches applied in the domain of object detection and tracking in videos. They usually apply expectation-maximization algorithm to estimate parameters from training set but the performance of Gaussian Mixture model gets deteriorated because parameters are bounded by uncertainty. In order to handle such type of uncertainties, type-2 fuzzy Mixture of Gaussian models [57,59,60] were developed. Type-2 fuzzy based Gaussian mixture modeling solutions has been very popular to handle high degrees of uncertainty as required in applications like surveillance. For object tracking problem, solutions based on fuzzy particle filter have clearly outplayed standard particle filter and Kalman filter based methods. Fuzzy particle filters [80,86,87] were developed to generate balance between system precision and computational cost in the particle filter framework. Divergence problem is regarded as the primary limitation with respect to Kalman filter which leads to limit their applicability for object tracking in videos. Therefore fuzzy inference system has to be integrated in Kalman filter [88] to solve divergence problem. Despite the progress made so far by fuzzy logic in numerous applications such as traffic/video surveillance, traffic monitoring, etc more efforts are required to apply fuzzy logic in highly uncertain situation for object detection and tracking applications such as faster detection of crime. Research is needed to reduce computational overheads associated with fuzzy logic and development of fuzzy datasets for object detection and tracking in videos. Another field that could also inspire studies in fuzzy logic is automatic generation of rules and membership functions from input data for fuzzy inference engines as used by researchers [93,94]. Although object detection based studies are not available but studies in domain of electrical and computer networks are available from where inspirations could be drawn to benefit studies in field of object detection and tracking.

5. Evolutionary algorithm based approaches

Optimization is the process of devising an approximate solution to specific problem subject to certain constraints for which finding an exact solution is not possible or too time consuming. Based on the nature's analogy, several optimization approaches have been developed. Object detection and tracking applications utilizing evolutionary algorithm based solutions tend to formulate detection and tracking problem as an optimization problem and then use any nature inspired algorithm to solve it. There exist a fair number of such approaches for detecting objects in videos in the literature. Techniques like Genetic Algorithm [95], Particle Swarm optimization[96]; Simulated Annealing [97], etc. are natural candidates for such applications and have been widely used in proposing good solutions using some aspect of object detection and tracking problems. This section discusses growth of such studies in field of object detection and tracking.

5.1. Evolutionary algorithm based object detections

The use of an entropy based simulated annealed background subtraction method for target extraction and detection in videos was performed by Karasulu et al. [98]. They used concept of simulated annealing for performance optimization of traditional background subtraction method and framed detection of objects as an instance of p-median problem. The framed problem was then handled using image entropy and a new background subtraction algorithm based on simulated annealing was proposed. Their work showed that entropy based simulated annealed background subtraction method provides better results as compared to regular background subtraction algorithm(Please refer to C6 for pseudo code and detailed algorithm).This is rarest of work in which entropy based exploitations have been used with soft computing techniques for detection of objects. The study has demonstrated promising results as compared to standard background detection method when tested on CAVAIR dataset.

Kim et al. [99] proposed distributed genetic algorithms, in which population of chromosomes mapping to video frames were used to perform video segmentation. The initialization of chromosomes was based on segmentation results of the previous frame gener-

Table 3
Summary of Fuzzy Logic based methods for object detection and tracking in videos.

Problem Statement	Paper	Why Fuzzy Logic?	Approach	Strong/weak Points
Fuzzy Logic				
Difficulty in suppressing dynamic textures in background effectively	56	Fuzzy Color Histogram has ability to minimize color variations generated by background motion.	Fuzzy color Histogram	No estimation of parameters was required.
Gaussian Mixture models (GMM) have limited capability for modeling uncertainty.	57, 59, 60	Type-2 fuzzy GMM offers capability to handle higher dimensions of uncertainty resulting in smoother classification result	Type-2 fuzzy GMM	The proposed approach was more robust for rapid variation of background.
Type-1 fuzzy system cannot handle uncertainties using membership functions and lead to misclassification outline	61	Interval Type-2 Fuzzy logic system can deal with high uncertainties and reduce misclassification of extracted silhouette.	Interval Type-2 Fuzzy logic	Type-2 Fuzzy logic system handled part of background challenges and extending it for automatic learning made it suitable for real time dynamic environments.
Critical situations confuse belongings of pixel to background or foreground.	62	Sugeno integeral is applied in foreground detection for better classification of pixels.	Fuzzy Integral	The limitation of Sugeno integral was that it assumes measurement scale to be ordinal.
Uncertainty in classification of pixels and imprecision in localization of moving object causes false classifications.	63, 64, 65	Choquet integeral can be applied for foreground detection and adaptive fuzzy background maintenance approach may help in achieving smoother classification	Fuzzy Integral	The fuzzy adaptive maintenance rule provided a means of handling uncertainty in a natural way.
Conventional Background subtraction algorithm cannot handle uncertainties of environment.	68	Fuzzy BS algorithm using choquet integral can eliminate uncertainties present in traditional BS.	Fuzzy BS with Choquet integral.	The proposed scheme effectively improved object detection performance.
Dynamic and challenging background conditions causes difficulty in detecting moving object in videos.	69	Sugeno and Choquet integrals were incorporated to model dynamic environment.	Low rank weightage and fuzzy aggregated multi-feature similarity	The scheme worked well for dynamic environment but the computation cost was high.
Issues like occlusion, variations in pose, in and out of plane rotation and poor illumination accounts for problem in object detection.	71	Fuzzy component based geometrical model can acknowledge situations like occlusions and poor illuminations	Fuzzy classifiers	Impressive results were obtained for specific class of objects i.e. face and can be extended to any other class of objects such as pedestrians and vehicles.
Performance of system deteriorates due to external environment conditions such as rapid illumination changes, complex background, etc.	70	False detection pertaining to critical situations can be removed by using background subtraction approach based on fuzzy color difference histogram.	Fuzzy color difference histogram.	Attempted to reduce effect of intensity variation due to fake motion and rapid illumination changes in background.
Recovery from problems such as moving cast shadows that causes object shape distortion, object merging and object loss is a big challenge.	73	A fuzzy-based rule system for tracking people using simple tracking algorithms can be used to prevent errors caused by such critical situations.	Fuzzy classifiers	The presented approach achieved better accuracy rates than traditional shape based and motion based methods.
Edge-based classification in real-world video sequences is challenging problem under changing lighting condition.	74	Fuzzy rule-based classification can be combined with changing region detection scheme to provide accurate detection of such regions in outdoor scenes.	Fuzzy classifiers	The classification of video sequences was limited only to buildings captured under natural outdoor lighting conditions.
The insufficiency of traditional fuzzy object classifiers causes uncertainty in the classification tasks.	75	Two-stage behavioral fuzzy classifier is applied to progressively discriminate among pre-defined semantic classes.	Fuzzy classifiers	Scheme was able to model uncertainty in features data and perform inference using fuzzy rules to achieve better classification.
Difficult to accurately detect and track object in video sequences in degrading lights.	76	The algorithm can perform classification based on fuzzy logic to identify moving regions that correspond to unexpected objects in indoor scenes.	Fuzzy classifiers	The limitation of present approach was that it was focused only for indoor video surveillance.
Most of conventional detection methods are not cost effective due to complex computation involved.	77	A fuzzy rule-based reasoning system can work as classifier to adopt statistical information from set of audio-visual features as crisp input and produce semantic concepts corresponding to occurred events.	Fuzzy rule based reasoning approach	The method kept down cost due to lack of need to perform complex computations. One of the main limitations of the proposed method was its confinement to soccer sport domain.
Early human action detection is a challenging task because of issues such as uncertainty, vagueness and impreciseness.	78	Hybridization of fuzzy logic and computer vision can help to achieve early detection of human action.	Computer vision and fuzzy set theory	Hybridization of fuzzy logic and computer vision led to greater accuracy of early detection of human action.
Fuzzy Tracking				
Particle filter(PF) suffers from tradeoff between accuracy and computational cost.	80	Fuzzy particle filter can handle system complexity by employing low number of particles and can provide robustness.	Fuzzy Particle Filter	The fuzzy particle filter has achieved better accuracy with low computational cost as compared to Kalman filters.
Varying illumination changes hamper precision of tracked object.	81	Color feature association and fuzzy c-means mapping can be used to improve precision and reduce computation time	Fuzzy Particle Filter with c-means mapping	Better tracking with reduced computation time was observed.
Tracking on PTZ camera results in irregular response time, low or irregular frame rate and changing field of view.	82	Geometric and appearance features can be used to weigh particle filter samples for handling large displacement of objects and irregular frame rate.	Fuzzy Particle Filter with optical flow	Proposed scheme was able to handle short term occlusions also based on certain assumption.

Table 3 (*Continued*)

Problem Statement	Paper	Why Fuzzy Logic?	Approach	Strong/weak Points
Rapid illumination changes results in transformation and deformation of target.	83	Monte Carlo algorithm can be employed with fuzzy particle filter to extract target features.	Fuzzy PF with Monte Carlo algorithm	Re-sampling of weight results in reducing number of required particles and increase robustness in illumination changes.
Binary logic does not permit enhancement of analysis of moving objects which directly affects tracking performance.	84	Video sequence motion tracking method by fuzzification can be used to tackle uncertainty of permanency values due to image noise.	Fuzzy logic	Integration of logic based decision rules effectively reduced complexity of approach and enabled grouping of spatial objects into coherent classes.
Tracking performance in optical flow is affected because of coupling between tracking and predictive algorithms.	85	Fuzzy control system is used to solve coupling problem between the different velocities in an efficient manner.	Fuzzy Logic	Object occlusion problem in video sequence was resolved by proposed approach.
Unexpected and improper noise variances occur due to change of state of tracked object.	86,87	Fuzzy particle filter with fuzzy noises can achieve better tracking performance by managing uncertainty in various noise models.	Fuzzy Particle Filter	Fuzzy particle filter was found to reduce tracking performance which takes a long time to estimate states than the conventional particle filter.
Conventional Kalman filter algorithm fails in case of stereo visual tracking.	88	Fuzzy controlled Kalman filter can solve stereo visual tracking problem by including fuzzy inference for noisy data	Fuzzy Kalman filter	System improved tracking performance to great extent than classical Kalman filter in stereo visual tracking problem.
Random noise due to dynamic background is major challenge in optical field that affects tracking performance.	91	Fuzzy hostility index is used to handle incoherent optical flow field containing random noises in efficient manner.	Fuzzy optical field	The proposed approach was efficient for real time tracking of high speed object in videos.
The traditional blob tracker fails in case of situations such as background clutter.	92	Combination of three-dimensional feature and a fuzzy membership distribution technique can be used to boost tracking performance.	Fuzzy logic	The results show that proposed approach achieves better tracking performance than traditional blob tracker.

ated randomly and consisted of a label and feature vector. These operations were done on locally distributed subgroups of chromosomes rather than entire populations. The fitness function was defined as difference between estimated color vector and actual color vector measured at a location in the image and chromosomes classified into stable and unstable categories. When an input frame is received, the stable chromosomes correspond to background and unstable chromosomes correspond to moving targets. The classification of chromosomes was done on the basis of fitness function values with chromosomes having low values classified as unstable. The mutation rate depended upon the degree of motion in a video and was instrumental in reducing computational cost between two successive frames. The work evidently showed effectiveness of proposed approach in terms of its speed and segmentation quality. Another method based on genetic algorithms and spatiotemporal restoration was presented by Kim et al. [100] for automatic object detection in videos. In this approach genetic algorithm was used to detect region of interest automatically. Background regions which were occluded by detected regions were restored using spatiotemporal restoration. Furthermore, input sequence was classified according to object motions and background frame difference to improve efficiency of restoration. The combination of spatial and temporal information in genetic algorithms was proposed by Kima et al. [101] for video segmentation. A genetic algorithm based method was applied for spatial segmentation to detect accurate boundaries and change detection mask was produced by temporal segmentation to determine foreground and background regions.

GA-based method with new evolution mechanism was proposed for segmentation to improve the performance. The segmentation is performed by individuals, initialized using the segmentation results of the previous frame, that independently evolve using distributed genetic algorithms (DGAs). Initially individuals were assigned random values with same evolution probabilities. The subsequent frames individuals were initiated from the segmentation results of the previous frame which were classified as stable and unstable individuals according to their evolution probabilities. The CDM produced by temporal segmentation was initially calculated by adaptively thresholding the intensity difference between two consecutive frames. This was refined and compensated by historical information for enhancement of coherent segmentation results of moving object. The objects were extracted using combination of spatial segmentation result and CDM. The objects were tracked using natural correspondence and post-processing. The natural correspondence means those regions where objects will maintain the same label throughout video sequence and temporal linkage was built using natural correspondence to the same objects in the subsequent frames. If such correspondence does not exist further post-processing was done which considered the size and location information of the object to be used. Kim et al. [102] used genetic algorithm for generation of effective and clear background. The work was targeted at removing errors caused due to illumination changes and noise in sequential images. Clear background was generated using suitability of changing regions. Evolution to next generation was done using chromosomes that were coded using edges and intensity values of the image to avoid effects caused by illumination changes. Comparison of results with conventional methods demonstrated effectiveness of proposed technique.

Input:- **Input image.**
Output:- **Threshold value for Simulated annealing based background subtraction method.**
Variables:- T_0: initial temperature, T_F: Final temperature, L_M: Markov chain length, a: current step, s:solution
Step1: The background image was initialized and converted to grayscale.
Step2: The histogram of the image was normalized and cumulative distribution function was computed for this histogram.
Step 3: Entropy method for grayscale image thresholding was employed.
(i) Region $A(x_1,y_1)$was assigned for dark pixel entropy.
(ii) Region $B(x_2,y_2)$ was assigned for bright pixel entropy.
Step 4: The foremost solution s_a^{old} was taken as location in Region $A(x_1,y_1)$ and Region $B(x_2,y_2)$ for current histogram entropies.
Step 5: The following parameters were initialized $T_0 = 1.000$ $T_F = 0.00001$, $L_M = 128$.
Step 6: while $L_M \neq 128$ and $T_F \neq 0.00001$ do
Step 7: s_a^{new} is generated using random number generator.
Step 8: Compute objective function $\Delta cost = cost(s_a^{new}) - cost(s_a^{old})$.
Step 9: Determine energy alteration $\Delta E = E(s_a^{new}) - E(s_a^{old})$.
Step 10: if $\Delta E \leq 0$ then $s_a^{old} = s_a^{new}$
else
Random variable is selected X~ uniform(0,1) and if P(Ta)>X then $s_a^{old} = s_a^{new}$.
endif
Step 11: Objective function was recomputed $\Delta cost = cost(s_a^{new}) - cost(s_a^{old})$.
Step 12: The location of RegionA(x_1,y_1) RegionB(x_2,y_2) was reassigned.
end loop
Step 13: The determined threshold was assigned for Simulated annealing based background subtraction method.

C6:-Pseudo Code for utilization of Simulated Annealing for Object Detection [98]

Nature inspired algorithms are also used for object detection and tracking for handling variety of issues prevalent in this field. Improved particle swarm optimization algorithm was used for moving object detection in videos by Yu et al. [103]. Improved-PSO utilized spatial segmentation integrated with Gaussian Mixture model for temporal segmentation in videos. Fusion of brightness and color information was done to improve PSO performance and adapt it to real time videos. The method generated optimal threshold value using particle swarm optimization. In order to improve competency of basic Particle Swarm Optimization approach for object detection, Cagnoni et al. [104] proposed modified basic Particle Swarm Optimization approach for two image analysis task. The modifications were done in two ways: At first forceful division of swarms into sub swarms was done and secondly particles dispersion was favored through regions of interest. k-means PSO was applied to achieve division of swarms into sub swarms and second goal was achieved by updating fitness function and velocity equation. It resulted in improvement of speed for image segmentation process and handling of complex background. A comparison of PSO and GA to propose an optimal neural architecture for detection of objects was presented in [192].

Manisha et al. [217] proposed object detection and tracking method based on fuzzy 2-partition entropy and Big Bang–Big Crunch Optimization (BBBCO) for optimizing the performance of standard background subtraction algorihtm. This was done by applying the concept of fuzzy 2-partition entropy and framing the problem of threshold detection as an optimization problem(Please refer to C7 for pseudo code and detailed algorithm). They used concepts of Big Bang Big Crunch Optimization to optimize detection of effective threshold value for a frame in consideration. BBBCO was also used to extract several other parameters for background modeling and detection of object regions. The detected regions were further tracked by Kalman Filter. The performance of the proposed appraoch was evaluated using benchmark datasets. The confirmation experiments results prove that proposed approach has better performance than the other methods such as traditional background subtraction method, simulated annealing based background subtraction method etc.

Input: - Frames from video
Output: - Detected object with bounding boxes
Step 1 : Read a video sequence and extract frame information and other video parameters and Initialization background as first frame
Step 2: Generate an optimized threshold value along with other parameters for current frame using BBBCO based fuzzy 2-partition entropy BS algorithm
(i)Initialize BBBCO parameters and population randomly in accordance to population size and number of parameters
(ii) Find probability distribution of gray level values in frame
(iii) For each gray level value perform fuzzification of foreground and background
(iv) Find pixel-wise fuzzy entropy for foreground and background regions and total fuzzy entropy of frame.
(v) Calculate fitness of individuals and find center of mass.
(vi) Reorder all feasible solution candidates according to fitness.
(vii) Calculate center of mass for current population of candidates and generate new population of candidates
(viii) Save the best solution and extract values of parameters like threshold, new learning rate etc.
Step 3: Evaluation: Perform evaluation and update background and foreground from extracted parameters
Step 4: Perform back ground modeling using Running Average method using updated value of running rate returned by BBBCO procedure.
Step 5: Generate binary frame using threshold values returned by BBBCO.
Step 6: Apply Morphological Processing for enhancing detection of objects
Step 7: Perform blob analysis to get objects of interest with bounding boxes and other parameters.
Step 7: Tracking of detected objects
(i) Compute centroids of blobs of detected objects and apply minimum bounding boxes to regions of object.
(ii) Input these regions to multi-object kalman filter based tracker.
Step 8: Plotting with bounding boxes: Find contours and draw bounding boxes on detected objects.
Step 9: Go to step 2.

C7:-Pseudo Code for utilization of Big Bang Big Crunch Optimization for Object Detection [217]

Evolution based algorithm based studies have been exploited by relatively large number of researchers in field of medicine. A study based on genetic algorithm (GA) was proposed by Bhuvaneswari et al. [105] for detection of cancer in lungs. Another GA based method for detection of tumor in brain was proposed by Chandra et al. [106]. The study has used genetic algorithms to accept enhanced, soft thresholded images as input. GA uses standard crossover and mutation operators to perform segmentation of MR images. Automatic thyroid disease detection was proposed by Liu et al. [107] using fuzzy based k-nearest neighbor (FKNN) classifier. The contribution of study was to explore capabilities of FKNN by automatically determining the neighborhood size k and fuzzy strength parameter m for thyroid disease diagnosis. In order to achieve this goal, continuous PSO was employed to dynamically conduct parameter optimization for FKNN classifier. The performance of proposed approach was evaluated using thyroid disease dataset and achieved accuracy of 98.82% compared to traditional approaches. More Recent attempt was made by Lahmir [108] for detection of Glioma using particle swarm optimization based on multi-fractal features of segmented brain MRI. The proposed system segments a MRI image by three variants of PSO namely classical PSO, Darwinian particle swarm optimization (DPSO), or fractional-order DPSO (FODPSO). MRI pixel set was seen as a problem space, and PSO meta-heuristic seek to find a pixel partitioning that maximizes between-class variance of the distribution of intensity levels. Classification of pixels has been done by trained SVM. Sweety et al. [109] used particle swarm optimization (PSO) and decision tree based classification for detection of Alzheimer's disease. The proposed scheme was implemented in three phases of feature selection, feature reduction and classification. Role of PSO was to generate reduced number of features so that a low computational classification tree could be generated.

5.2. Evolutionary algorithm based object tracking

In order to track targets of variable size in videos of low frame rate, Boroujeni et al. [110] proposed multi-objective co-evolution genetic algorithm. The merit of work was its robustness against temporal resolution changes and its application on videos with frame rate of 5 fps rather than 15 or more fps. Sample impoverishment problem in particle filter [111,112] was solved by Han et al. [113] using an evolutionary particle filter with immune genetic algorithm (IGA). IGA was used for target tracking by adding IGA before re-sampling process to increase particle diversity. The immune system brought out number of new antibodies largely through crossover and mutation operators. Visual target was selected in the initial frames manually. The color distribution was then used to establish similarity measure between sample and reference target. New particle set was calculated using random drift model and particle weights were calculated. The state estimation of the target was estimated and crossover and mutation was performed for evolving new particles. Re-sampling of particles was done to obtain new particles iteratively until targets were accurately tracked.

It was found that proposed particle filter integrated with immune genetic algorithm outperformed conventional particle filter.

Particle swarm optimization algorithm that used multiple interactive swarms to track various pedestrians sequences in a crowd was proposed by Thida et al. [114] as shown in Fig. 8. In order to track multiple targets in a crowded scene, motion prediction through particles and swarms diversification was introduced into multi-swarm particle swarm optimization. Swarms were initialized for every new target entering a scene and each swarm had 'n' particles where best state of the pedestrian was represented by its centroid location, width and height of the bounding boxes. Gaussian distribution was used for sampling at the beginning of PSO iteration in every frame. The target motion was predicted based on its personal information and movement of other members of social group. Two targets were considered as neighbors if their similarity score was greater than predefined threshold. The method employed histograms of oriented gradients (HOG) to obtain detections in each frame. Then matching score based on the spatial proximity, size and the appearance similarity was used to calculate difference between detection and current state of the target to be tracked. Finally, maximum matching score was used to find the exact match of the current state of the target to be tracked. Zhang et al. [115] adopted sequential particle swarm optimization framework for object tracking in videos by incorporating temporal continuity into conventional particle swarm optimization algorithm. In PSO based method, tracked object was localized using rectangular window for each frame of the video and affine image warping was employed to track object between successive frames. The target appearance was modeled using spatial constraint Mixture of Gaussian and online EM algorithm was used to estimate the parameters. Another work on tracking based on Swarm Intelligence Metaphor was detailed by Canalis et al. [116] in which a predator-prey metaphor was defined. The feature vector of prey consisted of gradient and color information. A template was defined as prey pixels' sample which will supply scent of preys to be tracked to swarm of predator particles. Each predator particle will track its prey using prey scent similarity principle. Authors argued that method is robust even if the object is deformed or image is cluttered since swarms follows light intensity independent colors. The method was found to be computationally less intensive but accurate tracking of small sized objects was an issue. To achieve efficient tracking, Zhang et al. [117] proposed a method which integrates PSO-based searching method and Bayes'-based probability classifier algorithm. Changes in scene are identified by particle swarm optimization and Bayes' based probability algorithm was used to provide an estimation of best candidate in the scene. In this framework, window based PSO was employed in which window parameters such as central point of the particle, width and height of the window and orientation of the particle were used for tracking. Different fitness functions were used for evaluation of these particles. The proposed algorithm considered object features and object motion estimation to speed up searching procedure. The system was tested in dynamic environments however number of tests conducted were very small and false detections were observed when object gets out of sight.

Yang et al. [118] used PSO approach to optimize unscented particle filter based tracking algorithm to prevent impoverishment problem. The algorithm produced a proposal distribution by unscented Kalman Filter method to improve sampling distribution function than conventional particle filter algorithm. Incorporating new observations in sampling process tend to optimize distribution of particles and problem of particle loosing diversity was avoided. The particles tend to converge to high likelihood regions which led to improvement in estimation accuracy. Problem of particle impoverishment was also targeted by Zhou et al. in [119] in which particle swarm optimized algorithm was integrated into particle filter framework. PSO method drove particles to explore solution space and search for the best-fit state estimation. Particle filter states were modeled using parameters such as (x,y, sx,sy) where x,y represents the target location, sx and sy represents the velocity of the target in x y direction. The moving target was represented using every particle swarm in this space. Initialization of best solution was done followed by calculation of global best solution. Sequential Monte Carlo method within Bayesian framework was utilized by particle filter to approximate probability density via a set of weighted samples. Bhattacharayya distance measure was used to measure similarity between discrete probability distributions. An

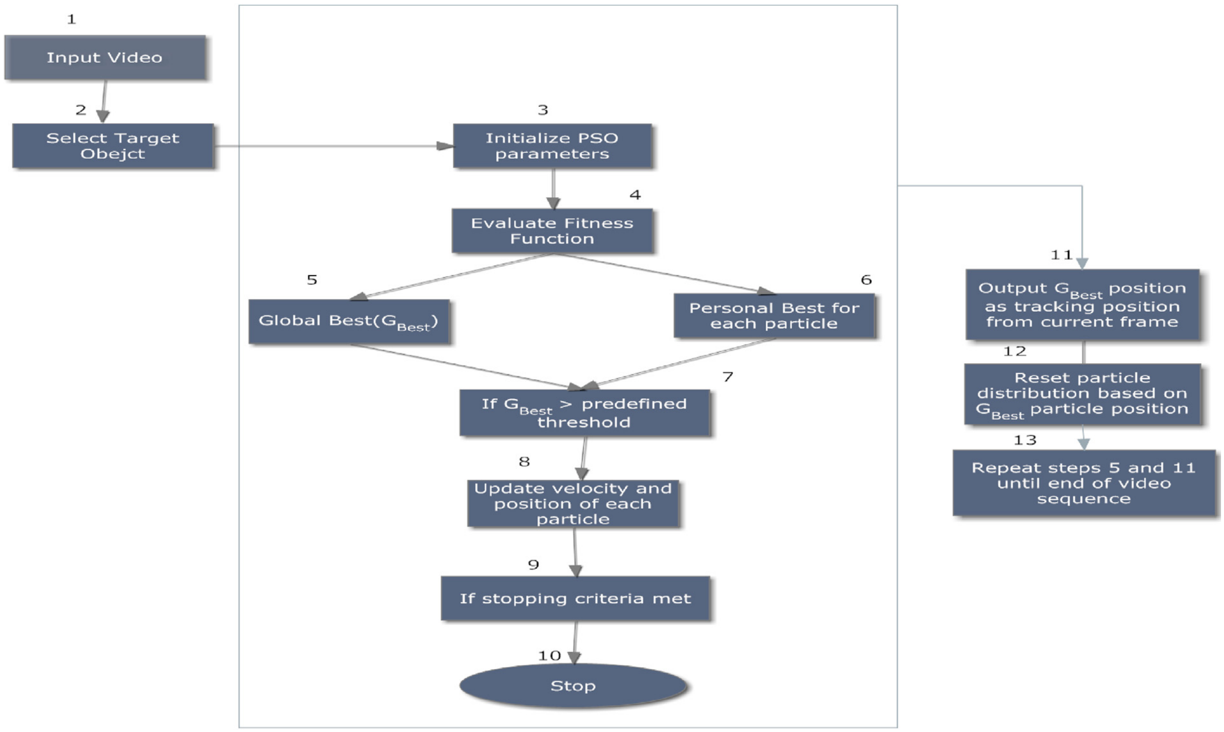

Fig. 8. Steps involved in object tracking using Particle swarm optimization.

adaptive discrete swarm optimization for video tracking framework was proposed by Bae et al. [120] in which object being tracked was represented using rectangle around it. The object was defined using four dimensional variable (x_0, y_0, w_0(width), h_0(height)) where x_0, y_0 represents 2D co-ordinates of the center point of the particle. In the next step, particle swarm was applied within a four dimensional feature space. The particles evolve to compute its position based on global best among all particles and an adaptive parameter. HSV histogram computed by image extracted from search window of target object was used to build a feature model for target object. The initialization of particle swarm was done around the target object within the search space. In every frame, the particle's current position was set as individual best and the particle with highest fitness value as global best during beginning of each iteration. The movement of particles was limited to sub-search space to find an optimal match of the target. In next successive frame particle distribution was updated based on the global best position computed from the last frame. In case of disappearance of target object or occlusion, the particles updated their search strategies to recapture target object. Compared to traditional PSO and GA algorithm, adaptive discrete swarm optimization achieves high accuracy and faster speed for tracking objects which were modeled in a high-dimensional feature space. Biogeography-based optimization was used for localizing a foreground object with respect to its background by Wang et al. [121]. Initially a set of local, regional and global salient feature maps were obtained and managed. In the next step, an optimal weight vector was evaluated using improved variants of biogeography-based optimization algorithm which was applied to link these feature maps into one saliency map. Finally, salient objects were identified successfully from the image background. The system had been shown to be more effective than other state-of-art methods.

In the past decades cost of sensors had decreased dramatically which has led to use of high dynamic range of infrared sensors in situations such as all weather and night-vision surveillance. But a single cue is insufficient to use with these entire environment conditions so multiple cues have been combined to obtain better tracking performance by Zhang et al. [122]. The proposed approach applied a multi-cue based PSO algorithm to make PF particles advance in direction of high likelihood area and found optimal multi-cue weight configuration in the update stage of PF framework. The equal weighted particles were generated according to prior density as the first step. The proposed approach used intensity and gradient as two cues for feature representation. The intensity and gradient histogram were computed for each particle and bhattacharya co-efficient was used to measure the similarity between histograms. Finally, the feature weight was yielded by calculating the pixel distance between predicted position and real target position expressed in terms of discrimination of gradient and intensity cues and the particles were then propagated independently. Resampling was done based on the efficiency criteria and the particle diversities were enhanced using random propagation. At last the histogram model of the object was updated to capture pose and appearance variation of object during tracking process. The authors had tested performance of particle filter tracker using three different sets of infrared videos.

Another PSO based tracking method for real time tracking of objects was proposed by Tawaba et al. [123]. A multi-feature model was used to detect objects to improve tracking performance. Target to be tracked was exemplify using rectangular window centered at the center of the object and delineated with four parameters (centroid_x, centroid_y, width, height). Two feature vectors were used to describe target area. The foremost one calculated the target's gray level intensity properties and the other one is the gray level probability density function (PDF) of the object to be tracked. Each particle was initialized primarily using Similarity (SIM) fitness function but if SIM was not able to locate the object correctly then the BC coefficient was used to capture the location of the target. The target was located manually and then targets' mean and standard deviation was computed for the similarity measure and PDF for the Bhattacharya coefficient (BC) fitness function. Edge enhancement detection algorithm was used to adjust the reference target parameters such as boundary and centroid for better results. The model amalgamates effects of various object cases into a single cost function.

The use of Particle swarm optimization and adaptive block-matching algorithm has been demonstrated by Kader et al. [124] for efficient eye detection and tracking system. The researchers applied PSO-based deformable multiple template matching technique to compute best candidate for center of eyes within a frame of the video sequence. The block-matching algorithm with adaptive search area was employed for eye tracking. The proposed object detection and tracking approach was successful in providing significant improvement in performance than deformable template matching based methods such as genetic and PSO. Subsequently

Zhao et al. [125] also applied Particle swarm optimization algorithm to improve performance of Particle filter in heavily cluttered and complex scenes. The authors presented a re-sampling method based on Particle Swarm Optimization to compute high likelihood area to distribute particles in efficient manner. In this framework the state of the particle was expressed with object position. The blend of color histogram (HC) and histogram of orientations gradient (HOGC) was used as observation model. The HOG features were extracted on each gray level image window. PSO was used for searching new sample area around the position of the object. Ten particles were sampled around t-1 time step and Bhattacharya coefficient was used to measure the similarity between the candidates and object model. The method was demonstrated to work effectively and efficiently on real-tracking scenarios than traditional particle filter.

5.3. Discussion

Table 4 depicts collection of various evolutionary based algorithms used for object detection and tracking for videos. Simulated annealing has been exploited to propose an entropy based object detection method in [98] however scheme suffers from computational delay. The advancement in soft computing based approaches for object detection and tracking in videos has lead to use of Genetic algorithms and particle swarm optimization in handling numerous computer vision and image processing challenges. Both PSO and GA have clearly outclassed all other evolutionary techniques. This may be attributed to the fact that both PSO and GA provide quick convergence and can overcome local minima quickly. In the past, they had already proven their efficiency when tested against well known algorithms. On investigating the results in current domain, it was found that Genetic algorithms and particle swarm optimization also perform better than other evolutionary approaches in terms of accuracy [104,113,119,124]. Another important observation is that because of simplicity, easy implementation and quick convergence of PSO and GA; these are more suitable for complex optimization problem like medicine. Therefore capability of both evolutionary approaches is emerging as important aspect for object detection and tracking in videos and applications like medicine. The literature has provided very little head to head comparisons among these approaches but when investigated deeply it was found that PSO scores over Genetic algorithms [114,120,192]. It was observed that for an equal number of particles and equal number of iterations, PSO gets an edge over genetic algorithms. For object detection methods, PSO provides better results while measured in terms of parameters like standard deviation, mean and best recognitions. PSO gives better results for tracking also when measured in terms of parameters like average processing time per frame, error positions per frame etc. This may be because of advantages of PSO over GA, as PSO allows greater diversity and exploration over a single population (which with elitism would only be a population of pbests). This is different from GA in which algorithm has to undergo sequence of operations like selection, mutation, crossover and best solution is usually not preserved in different evolutions. The momentum effects on particle movement can allow faster convergence (e.g. when a particle is moving in the direction of a gradient) and more variety/diversity in search trajectories which is not the case with GA or similar evolution based methods. Future directions for evolutionary studies can seek inspiration from new variants of PSO like adaptive PSO and adaptive genetic algorithms. Adaptive nature of such techniques allows position updates according to fitness values of particles and makes each particle's searching ability more adaptive. Moreover size of search area can also be dynamically reduced or enlarged. Exploitation and use of some recent optimization approaches can be further explored for designing some real time applications. One such study is available at [121] in which biogeography based optimization was used but an honest attempt is still desired in this direction.

6. Hybrid approaches

Hybrid approaches uses combination of more than one soft computing approach to effectively detect or track object in video sequences. Efficacies of multiple approaches are combined to overcome limitations of one approach and design effective decisions using combination. Research in hybrid approaches for object detection in videos had always been targeted to device intelligent techniques to enhance functionality of object detection approaches. Therefore discovery of hybrid approaches is actually a big achievement and some of work carried out by researchers based on these approaches is summarized in this section.

Neuro-fuzzy approach was applied by Maddalena et al. [126] for detecting moving objects in video sequence. Their paper has addressed issues related to uncertainty in detection process which was caused by background maintenance phase of Background subtraction approach. The logic behind proposed technique was to implement self organizing map approach to detect moving object based on background model. Fuzzy learning factors were introduced following classification of the pixel in update procedure of background model of Self Organizing Background Subtraction. The experimental results on real color video sequence showed promising results and redirected researcher's attention towards hybrid approaches of object detection in videos. In fact, introducing fuzzy approach in maintenance phase of object detection procedure in videos improved accuracy of corresponding crisp moving object detection procedure. An approach based on similar concepts was presented by Murgia et al. in [127]. The authors had tried to apply neuro-fuzzy approach to extract moving object in videos having dynamic backgrounds. The self-organized mapping (SOM) from neural network was used so that system was able to adapt to different changing environment conditions such as shadowing. Further, fuzzy inference was applied to deal with dynamic changes in video by regulating parameters of the SOM-based background model according to state of moving objects. Self organizing background subtraction model by incorporating spatial coherence and fuzzy based approach was presented by Maddalena et al. [128] to handle uncertainties and imprecision. Spatial coherence was introduced into background subtraction to guarantee more robustness against false detection and fuzzy functions were used in learning step to improve moving object detection accuracy. The term spatial coherence was defined in terms of intensity difference between locally contiguous pixels and small intensity difference shown by neighboring pixels was considered coherent. A neuron map consisting of n^*n vectors, for each pixel, was represented in HSV color space and current frame was subtracted from the background model to determine existence of weight vector matching current pixel weight vector. The pixel was classified as foreground if no acceptable weight vector exists otherwise it was classified as background.

Table 4
Summary of evolutionary based methods for object detection and tracking in videos.

Problem Statement	Papers	Why Optimization Based Approach?	Approach	Strong/Weak points
Evolutionary Based Approaches				
Complex backgrounds, dynamic lighting conditions lead to ineffective object outcomes.	98	Entropy based simulated annealed background subtraction method is used for generating optimal threshold because of its effective exploitation within search space	Entropy based simulated annealed background subtraction method	The major limitation of algorithm was slow convergence speed.
Slow convergence speed and low search efficiency of classical genetic algorithm reduces segmentation performance in videos.	99	Distributed Genetic Algorithm can effectively utilize temporal correlation between successive frames.	Distributed Genetic Algorithm	Parallel exploration of search space resulted in faster convergence but segmentation results required further improvement
Conventional approaches are not able to perform occluded background recovery.	100	Genetic Algorithm with spatiotemporal restoration was able to perform occluded background recovery for better object detection.	Genetic Algorithm	This method can be adapted to stationary/moving object on stationary background in videos.
Traditional approaches for segmentation suffer from over segmentation and intensive computational cost	101	Spatiotemporal segmentation can automatically decompose a video into objects making up scene.	Genetic Algorithms	The proposed approach was limited to video sequences with stationary background.
Generated background is affected by illumination changes.	102	Genetic Algorithm can use edges and intensity values of frames to avoid effects of illumination changes.	Genetic Algorithm	The proposed approach was capable of handling ghost and illumination change problems.
Conventional PSO has difficulty in managing sudden changes in light conditions.	103	Improved PSO supports dynamic updated segmentation and is independent of prescribed evaluation function.	Particle Swarm Optimization	Succession strategy has strong global Optimization ability which makes it suitable for real time processing.
Basic Particle Swarm Optimization is sensitive to noise and possibly misleading	104	Improved PSO achieves dynamic reconfiguration of swarms based on clustering techniques to improve classification performance.	Particle Swarm Optimization	Offered a solution towards low accuracy of conventional Particle Swarm Optimization approach.
Object tracking				
Particle filter methods cannot track objects having variable sizes within frames	110	Multi-objective co-evolution GA is developed to track variable size objects in low frame rate videos.	Genetic Algorithm	It can track far and variable size objects better even in low frame videos.
Particle degradation becomes inherent shortcoming of particle filter.	113	Immune genetic algorithm (IGA) is added before re-sampling process to increase particle diversity.	Particle filter with genetic Algorithm	The particle set optimized by IGA expressed true state of target, and number of meaningful particles increased significantly.
The standard PSO cannot address problem of tracking multiple targets.	114	Multiple target tracking can be achieved by incorporating motion prediction and social interaction in standard PSO approach.	Particle Swarm Optimization	Proposed method was able to track multiple targets in crowded scene with severe occlusion.
Conventional PSO fails to track objects in dynamic environment.	115	Sequential PSO offers more stable performance by incorporating temporal continuity into conventional PSO.	Sequential Particle Swarm Optimization	Sequential PSO based tracker is better both in terms of accuracy and efficiency than particle filter.
Difficult to track object with arbitrary shapes using region template tracking or using salient features	116	Pixel color and image information are utilized to form swarms which are tracked using combination of different rules.	Predator–Prey scheme based swarm intelligence	Proposed method worked with occluded objects but accuracy for small sized objects was an issue.
Using only object features in PSO is not efficient and less robust.	117	Object motion can also be taken into consideration using Haar like features to form cascade classifiers to find best match.	Adaptive particle swarm intelligence	Proposed method was able to track objects in indoor environments however false detections can be observed.
Serious degeneracy phenomenon can appear in simple PF because of impoverishment problem	118	Improved PF algorithm based on choosing a good proposal distribution to obtain better sequential importance sampling.	Particle Swarm Optimized Unscented Particle Filter	The solution is more accurate and robust and accuracy of video tracking is improved as perceived by visual results.
A large number of samples are required to properly approximate posterior density of state evolution in classical PF.	119	PSO can be incorporated into particle filter after prediction stage to reduce impact of impoverishment.	Particle swarm optimization embedded into generic particle filter.	Embedded PF effectively reduced size of sample set and improved accuracy and robustness of tracking system.
Traditional PSO sometimes fails in finding true global optimum in some situations.	120	More stable and fast tracking performance was obtained by applying Adaptive Discrete swarm optimization technique.	Discrete Swarm Optimization	The proposed method does not reflect comparison with state of the art method in terms of some specific parameters.
Detecting salient objects in an efficient and effective way is a great challenge.	121	Biogeography-based optimization is applied to combine features into one saliency map for labeling salient object from image background.	Biogeography-based optimization	The method outputs uniformly highlighted salient regions with well-defined boundaries.
Single cue for tracking is insufficient to deal with environmental conditions like nights or bad weather.	122	A multi-cue based PSO guided PF tracking framework is employed to deal with certain circumstances.	Particle swarm optimization (Particle filter (PF) tracking)	Multi-cue based convergence criterion helps avoiding local minima.
Real time environmental issues effects tracking performance.	123	Multi-feature PSO method was employed to increase tracking performance for real time applications.	Particle Swarm Optimization	Simple particle to particle interaction made it suitable for parallel and real time processing.
The efficiency of eye tracker reduces due to head movements, eye blinking, and occlusions.	124	PSO and adaptive block-matching algorithm helps in relaxing limitation caused by critical circumstances for eye trackers.	Particle swarm optimization and adaptive block-matching algorithm	Better accuracy and computational complexity than edge projections based method.
Particle filter looses performance in cluttered and complex scenes.	125	High likelihood area for distribution of particles can be found by resampling method.	Particle Swarm Optimization	The method was tested on real time scenarios and improvement in performance was observed.

A hybridized method based on multi-valued Neural Network with Gaussian mixture model algorithm was proposed by Luque et al. [129]. They had used Gaussian Mixture model method proposed by Stauffer et al. [130] in devising video segmentation algorithm and multi-valued Neural Network was then applied to improve segmentation results. Experiments were performed on publically available videos sequences on internet and visually satisfactory results were only shown. Gallego et al. [131] employed a recursive background modeling approach to dynamically model background in which membership functions were tuned employing Marquardt–Levenberg algorithm within a fuzzy system. Results of experiments showed importance and application of Marquardt–Levenberg algorithm integrated within a fuzzy system in terms of improved object detection in videos. Another object detection technique that used neuro-fuzzy approach is available at [132]. The scheme presented background subtraction model using neural self organizing map with fuzzy automatic threshold update to handle uncertainties such as illumination changes and slight shadow problems. In this framework, each video frame pixel was modeled using HSV color space such as $p(x,y,t)=[H,S,V]^T$ and to build background model, each pixel of the frame was mapped into neuron $W(x,y,t)$. Euclidean distance was computed to find whether a pixel is a part of moving target in the HSV color-space from the second frame. Difference Value Analysis (DVA) was done on various situations such as scene with drastic changes, with normal changes, with no changes using analysis of difference between each pixel and neuron. A scene analysis mechanism was employed to update learning rate values of Background subtraction model automatically and a scene with drastic changes updated the background model with input information and DVA. Furthermore, optical flow algorithm was applied to analyze dynamic regions to improve detection of objects further. A new fuzzy estimator named Takagi-Sugeno-Kang (TSK) type fuzzy method for background estimation was devised by Li et al. [133]. This method was hybridization of particle swarm optimization with fuzzy logic and kernel least mean square method was used to train fuzzy estimators. The proposed approach was not able to handle problem of changes in illumination and in cases where foreground objects remains present over long time.

A hybrid approach using genetic algorithm and deep convolutional neural networks (CNN) was proposed by PaulIjjina et al. [134] for recognition of human actions. Solutions generated by genetic algorithm were used to initialize weights of a convolutional neural network to reduce classification error. The convolutional neural network was trained using gradient descent algorithm during fitness evaluations of chromosomes. The convolutional layer weight mask and seed value used by random number generator were considered as GA chromosome employed for fully connected neural network initialization. The fitness function for the evaluation of GA chromosomes was CNN classification error. A single convolution mask was used in the initial convolution layer because of simplicity of pattern and masks were doubled in the succeeding layers. The binary decoded output of the classifier helped in determining the action labels. Tests were conducted on UCF50 dataset to demonstrate performance of suggested method. Another interesting hybrid approach for human action recognition was exposed by Ijjina et al. [135]. They proposed hybridization of deep neural network model by taking fusion of CNN classifier with binary decoded outputs. Action bank feature of a video was used as input to CNN classifier for action recognition. The input data was diversified by complementary action bank classifier and applying varying model initialization to generate different versions of the same model. An effective classifier fusion model was designed by leveraging high confidence of classifiers for correct prediction and variations of model. Fusion using maximum value was used as classifier model to ensure selection of correct class labels with evaluations being done on UCF50 dataset.

Vision-Based Neuro-Fuzzy approach was presented by Wu et al. [136] for adjacent lane detection and lateral vehicle distance measurement for a driver assistance system. To capture images of adjacent lane, camera was mounted at left side rear view mirror of the experimental vehicle and image processing techniques were employed for detection of lane lines; and lane-based transformations were explored for the candidates for lateral vehicle. A well-trained recurrent functional neuro-fuzzy network was used to estimate distances of lateral vehicles by computing the real distance of the detected target from the pixel distance of the image. The pixel distance was measured till vehicle was 30 m away and this was repeated for every two meter backwards. After learning, only four rules were framed in fuzzy and training process was continued for 500 iterations. The 9 video sequences captured when the vehicle was driving on Taiwan's highway were used to evaluate the presented approach. Another hierarchical neuro-fuzzy architecture was devised by Acampora et al. [137] for effective analysis of human behavior. The raw kinematic data obtained by tracking algorithm was analyzed by employing suitably trained time delay neural networks (TDNNs). The architecture of TDNN composed of input layer, two hidden layer and one output layer. The input layer comprises of three features such as horizontal position, vertical position and speed. The hidden layer consists of 15 neurons and output layer consists of three neurons to provide information about object target in terms of three behaviors: walking, stopping and running. The input layer was connected to second layer with delayed time approach and TDNN was trained using back-propagation paradigm. The fuzzy inference engines were used to identify a collection of refined and context depending behaviors. The input to the fuzzy inference system was lower layer information such as micro behavior, object distance and appearances of the system. Trapezoidal membership function was used for modeling human behavior analysis concept. Mamdani fuzzy rule system with max-min inference and center of gravity method for de-fuzzification was employed. The neuro-fuzzy network based application called HebbR2-Taffic for traffic analysis was presented by Cho et al. [138]. A nine feature value vector was extracted for representing features of the vehicles and this vector acted as input patterns to neural network for performing training and validation. The proposed approach was implemented as five layered network namely: input, condition, rule node, consequence, and output layer. Non fuzzy vectors were used for representing input and output layer. Condition; rule node and consequence layer were used for performing fuzzy inference. The input layer transfers the raw input to the condition layer which acted as singleton fuzzifier. The condition layer transformed raw input into fuzzy input using Gaussian membership function. The activation function of each node in the rule node layer performs T-norm function. S-norm function was performed by consequence layer by aggregating output of each node in the rule node layer. Modified center average de-fuzzifier was employed at the output layer to convert the output into crisp value. Hebbian–Mamdani rule reduction architecture was used for classifying and counting vehicles passing through a three or four lanes expressway, as a result of which well-trained network was able to perform vehicle counting and classification accurately.

Hybrid techniques have been successful in proposing some wonderful solutions for benefits of society. In this context, the recognition of Alzheimer's disease (AD) was attempted by Cho et al. [139] using genetic algorithms and artificial neural network approach. A feature pool including 88 spectral, 28 statistical and 2 nonlinear characteristics of EEG and 10 features of ERP was made. The combined GA/ANN was applied to find the dominant features automatically from this feature pool. The detected features were given as inputs to neural network. The recognition rate of ANN fed by this input was 81.9% for the untrained data set. Furthermore, Kim et al. [140] also proposed detection method of Alzheimer's disease using genetic algorithms and artificial neural network approach with single channel EEG recording however results were not promising. Invent of Magnetic resonance imaging (MRI) techniques provides rich information for clinical diagnosis and biomedical research. The diagnostic values of MRI were greatly enhanced by automated and accurate classification of the MR images and played an important role in detecting AD subjects from normal elder controls (NC). A hybrid Eigen brain based CAD system that can, not only detect AD from NC but also detect brain regions related to AD was proposed by Zhang et al. [141]. The study used Eigen brain feature for the first time on MRI images and inter class variance was established using novel approach. Support vector machine (SVM) was used for classification with sequential minimal optimization(SMO) to train SVM. The testing was performed on RBF and polynomial kernals with different scaling factors; interclass variance and Welch's t-test (WTT) was used to reduce redundant data. PSO was used to obtain best parameters like scaling factor of RBF, or the degree and soft margin constant of POL with additional task of optimization of error penalty. The results were demonstrated in terms of visual images and various statistical measures.

The use of hybridization of biogeography-based optimization (BBO) and particle swarm optimization (PSO), for detecting tumors from normal brains in MRI scanning was presented by Zhang et al. [142]. They applied wavelet entropy for extracting features from brain MRI images which were fed to feed forward neural network for classification. The training of feed forward network was regarded as an optimization problem in which variables were considered as combination of the connection weight matrix between input and hidden layers, and between hidden and output layers. The training was performed with hybridization of BBO and PSO (HBP) so that exploration ability of BBO and exploitation ability of PSO could be combined. The whole population was divided into two with one implementing PSO and other BBO. A new offspring was produced from gbest of PSO and elite of BBO using a crossover operation. The hybridized approach with an accuracy of 99.49% with only 6 features outperformed other state-of-the art methods. Hybridization of PSO and artificial bee colony optimization (ABC) has been done to design CAD tool for differentiation of normal brains from abnormal brains with PSO and artificial bee colony optimization [143]. The proposed method used translation invariant stationary wavelet transform (SWT) to extract features from MR brain images. To improve performance of classifier on query images, three novel feed forward neural network variants (FNN): IABAP-FNN, ABC-SPSO-FNN, and HPA-FNN based on three different hybridization methods of PSO and ABC were proposed. Evaluations of these three hybridization methods had also been done on three different datasets for comparative purposes. Recently, Zhang et al.[144] proposed another method for detection of Alzheimer's disease and mild cognitive impairment using 3D-Discrete Wavelet Transform (DWT). 3D-DWT was used to extract features from a 3D brain which were reduced using PCA effectively so that dimensionality of the data and subsequently computational cost of analyzing new data can be reduced. The paper implements a multiclass classification method using winner-takes-all Kernel Support Vector Machine trained by Particle Swarm Optimization with time-varying acceleration-coefficient (PSOTVAC). PSOTVAC was chosen to enhance global and local search capability by giving more weight on cognitive component and less weight on social component at the former stage; and giving less weight on cognitive component and more weight on social component in the latter stage. This method was successful in achieving accuracy up to 81.5% as compared to others whose accuracy were limited to 74%.

6.1. Discussion

Table 5 provides **s**ummarization of various studies based on hybrid approaches. The table describes problem undertaken by study and how hybrid approaches were able to effectively handle this problem. There is need of hybridization of various soft computing approaches in order to enhance performance and efficiency of object detection and tracking algorithms. The table indicates that a large number of neuro-fuzzy based studies are prevalent in literature. Combination of both had proven to be major choice among researchers for implementing solutions for proposed problem. The main reason for combing neural network and fuzzy system is that some limitations can possibly disappear by combining both approaches. For example, Fuzzy logic is able to represent uncertainty but designing of such system is very difficult because rule formation and extracting information from inference engine can be complicated. Furthermore, it is difficult to make fuzzy logic system generic as rules are fixed and difficult to change. The solution to this problem is provided by neural network because they have the ability to learn and generalize. Fuzzy logic provides good modeling of uncertainties which are persistent in working environment while neural network based systems had sound history in classification of pixels. Therefore, hybridization of neuro fuzzy approaches is beneficial to both strategies. Deep learning has proved to be efficient approach due to its powerful learning characteristics as compared to traditional approaches. In spite of efficient learning characteristics of deep architectures, straightforward optimization of deep architecture has been proven to be computationally expensive problem. Some insights in this direction has been made by researchers by combining deep learning concepts with genetic algorithms but such studies are still in growing stage [134]. This was due to efficiency of genetic algorithms to generate background effectively and powers of deep neural networks to make effective classifications. The obtained results were appreciable. The use of evolutionary algorithm was done to address challenges in training deep architecture. Therefore various evolutionary algorithms can be applied to optimize the feature, topology and weight selection process of neural network architectures. Instead of applying conventional approaches like background subtraction for background modeling in multimodal videos, fuzzy logic has been used which is capable of approximating estimation function even in presence of noise. But performance was still not desirable and therefore PSO has been used in combination with fuzzy to optimize parameters in fuzzy inference system to improve its performance [133]. Since PSO has the ability to fine tune parameters by finding near optimal solutions in complex search spaces therefore it is suitable candidate for incorporating a priori knowledge in design of fuzzy rule base system which can also be future focus of researchers. Because of ability to fine tune parameters, PSO has also been widely appreciated to offer reliable solutions in field of medicine by hybridizing it with BBO and ABC optimizations. Gaussian mixture model have been widely used for object detection in videos but it suffers from spurious detection of objects. This situation arises due to inability of the approach in establishing relation with the neighborhood of each pixel which is used to obtain these objects in motion. Therefore, Gaussian mixture model was extended to Multivalue neural network Gaussian mixture model which is capable of eliminating the limitations of conventional Gaussian mixture model. The present section has discussed about the reasons for the hybridization and degree to which one technique is hybridized with other technique. It can be concluded that there is no general rule for hybridizing two or more approaches together and choice lies with the designer of algorithm. However it can be inferred that for modeling uncertainty of background environment. Fuzzy logic has shown success where as neural network or deep learning based solutions are good at classification decisions. The future directions in hybridization

Table 5
Summary of hybrid approaches for object detection and tracking in videos.

Problem Statement	Papers	Why Hybrid Approach?	Approach	Strong/Weak points
Hybrid Approaches				
Conventional Background Subtraction method is unable to model uncertainties effectively.	126	SOM integrated with Fuzzy approach can effectively increase detection performance in complex environment.	Self Organizing Map with Fuzzy Logic	The fuzzy approach used in the background update phase helped to deal with sudden illumination changes and shadows.
Dynamic background, sudden illumination changes, etc. affects object detection and tracking.	127	Shadow elimination and dynamic background are handled using SOM and fuzzy can be used to mimic human behavior for automatic parameter adjustment in SOM.	Neuro-fuzzy approach	The proposed approach improved segmentation results.
Self organizing background subtraction (SOBS) was not able to handle imprecision and uncertainties.	128	SOBS was improved by integrating fuzzy function in learning step to overcome SOBS limitations	Neuro-fuzzy approach	Integrating fuzzy approach with SOBS method improved accuracy of moving object detection procedure.
Multi-valued neural network was proposed to correct some deficiencies of GMM Algorithm.	129	Integration of Multi-valued Neural Network with GMM can effectively improve classification process.	Multi-Valued Neural Network with GMM	The study didn't include parameter setting mechanism and results provided were only qualitative.
Determining parameters like membership functions and threshold for background modeling in fuzzy system is difficult.	131	Machine learning integrated with fuzzy enabled system to learn optimum parameters to improve overall system performance.	Neuro Fuzzy Approach	Integration helped in handling more uncertainty which led to less human intervention.
Uncertainties such as illumination changes and shadow pose serious problems in traditional background modeling algorithms.	132	Neural Self Organized Map integrated with fuzzy's automatic threshold update mechanism makes model robust to uncertainties present in state of art reported models.	Neuro-fuzzy approach	Thresholds and learning rates adapted to dynamic environment automatically.
Random noise due to sources of disturbance in a scene affects object detection and tracking performance.	133	TSK type fuzzy model may be used to filter noise in an efficient manner;	Adaptive Fuzzy using PSO and Kernel least mean square(KLMS)	The system provided relatively slower convergence.
Insufficiency of shallow neural network models and lack of efficient training algorithm effects human action recognition.	134	Hybridization of genetic algorithm and convolutional neural is applied to improve prediction of human actions in unconstrained videos.	Genetic algorithm and convolutional neural network	Smoother classification results were achieved using proposed approach.
The use of single classifier reduces the classification performance.	135	The overall classification performance is improved by using a hybrid deep neural network model using fusion of CNN classifier with binary decoded outputs.	Hybrid deep neural network	The multi-classifier approach is more effective for detection of human action in videos.
Difficulty of providing efficient driver assistance system by traditional approach.	136	The limitations of traditional method can be reduced by estimating distances of lateral vehicle using vision based neuro-fuzzy approach.	Neuro-fuzzy approach	The system provided benefits such as lane detection, vehicle detection, and distance measurement for lateral vehicle.
Difficult to develop human behavior analysis framework because of uncertainty and vagueness in human daily activities.	137	A time-delay neural network for analyzing micro-behaviors and fuzzy inference system for analyzing macro-behaviors can be used for human behavior modeling	Hierarchal Neuro-fuzzy approach	The neuro-fuzzy system offered high scalability and tolerance for uncertainties that characterize human behaviors.
The insufficiency of MLPs and RBF network to accurately model object because of over learning training data.	138	Neuro-fuzzy approach generalizes training data to infer output for similar data correctly.	Neuro-fuzzy approach	The blend of neuro-fuzzy network approach demonstrated good efficiency for classification and prediction problem.

will lead to testing of evolutionary studies on deep architecture for better and automatic feature selection, feature reduction that can help to reduce computational complexity of deep learning and aid in developing better solutions.

7. Other trends in object detection and tracking

Apart from traditional soft computing based methods, some of new concepts have been recently seen in the field of object detection and tracking. These can be classified as learning systems based on representation and machine learning, knowledge based, representation learning and expert system based solutionss. This section is devoted to discussions based on such new concepts.

7.1. Learning based systems

The concept of learning has originated from pattern recognition and computational learning theory. It can be proficiently used for development of computer applications that can autonomously learn and make predictions based on new data. Such studies behave differently from neural based learning methods in which weights are adapted while features are more or less fixed. The studies discussed in this section tend to behave more like feature extractors for helping classifier in training. In order to prevent overlapping with previously discussed approaches, only feature extraction methods and classifier training methods are considered. For classifier training, numerous methods such as Adaboost, Support Vector machines or their combinations have been used for object detection and tracking in videos. Hybrid models based on Adaboost algorithm and convolutional neural network has been proposed by Gao et al. [145] for counting of people based on detection of heads. The general idea behind use of Adaboost algorithm was to extract head regions for CNN which can significantly decrease classification time. The AdaBoost-SVM classifier has the capability to remove limitations of traditional cascade-Adaboost classifier and also reduce time consumption of support vector machines when applied to large-scale training sample sets. The research of Cheng et al. [146] applied AdaBoost-SVM classifier for pedestrian detection in videos. The proposed classifier can select AdaBoost classifier or SVM automatically for development of a cascade classifier with respect to training samples for improvement of classification performance. In a similar study, Guo et al.[147] has used AdaBoost-SVM classifier for pedestrian detection in videos. The proposed model operated in two stages; First stage deals with segmentation of pedestrian candidates from images using AdaBoost algorithm and cascading method. The second stage deals with evaluating whether a candidate is pedestrian or not. Support Vector machine was used in second stage for training a pedestrian classifier. Finally, the pedestrian detection method was evaluated against traditional single-stage classifier, such as AdaBoost based or SVM based classifier. It was found to be more effective than both. Recently studies based on use of support vector machines [148] have been gaining importance in the field of object detection and tracking in videos. Zhang et al. [149] proposed hybrid model for object tracking in videos where spatial constraints were exploited using regression model. The method applied three different support vector machines for developing a tracker and hybrid support vector machines were combined using a distributed strategy. The experiments on this technique indicated that proposed approach was better than state-of-the-art. An algorithm for object tracking in videos based on Support vector machine was proposed by Cyganek et al. [150] and [151]. They had used one-class SVM classifier with mean shift tracker. A feature vector was built from components of orthogonal IJK color space and structural tensor. In the devised method, the structural tensor was evaluated in pixel regions to gather information related to intensity variation. Output of SVM, after training using these features was softened and was inputted to mean shift tracker. According to the study in [152], multi-view learning framework was introduced using multiple support vector machines for object tracking in videos. The method was constructed by taking into consideration multiple view of features such as gray scale value, histogram of oriented gradients (HOG), and local binary pattern (LBP). These features were used to train the support vector machines. The update scheme based on subspace evolution strategy was introduced to learn the changes in the object and the scenario.

Representation learning [153] is a set of techniques in which feature learning helps to extract needful information easily during constructing classifiers or predictors. Authors had exploited deep learning, feature learning and clustering based methods under this category. Most literature under representation learning domain has utilized deep learning for object detection and tracking in videos. Su et al. [155] has used feature learning method based on sparse auto-encoder (SAE), to handle problem of depth feature extraction for human body detection in videos. SAE is unsupervised feature leaning methods which can avoid the labor-intensive and handcraft feature design. The proposed method used convolution neural network and pooling to decrease training complexity of SAE. Sliding window localization strategy was proposed in which histogram of depth was used to develop candidate detection window center. The method further exploited the relationship between human body height and depth to evaluate detection window size. Clustering based representation approach has been used by Jia et al. [156] for detection of anchor person in news video sequence. Initially, new videos were divided into shots and thereafter key frames were extracted from every shot for representation of news video. Further, statistical distance of Pearson's correlation coefficient was used to perform clustering of key frames and anchorperson shots was detected using anchor shot criterions.

7.2. Knowledge based system

We have seen some meaningful advances in the field of creation of knowledge bases for object detection and tracking system in videos in recent years. Knowledge based systems help to develop object detection and tracking systems in an intuitive manner using examples for extracting information. These have been covered under statistical and syntactic approaches. Statistical approaches include studies on Bayesian belief network and Hidden Markov models while syntactic approaches include studies on approximate reasoning of knowledge based system. Bayesian network has become a popular choice for tracking objects in videos because they have ability to model temporal relationship. A typical example of such system can be found in the work of Yao et al. [157] where multiple objects in videos were tracked based on Gaussian mixture and dynamic Bayesian network. They applied mixture of Gaussians with spatial prior for background segmentation and has used online approximation approach for updation of the model. The authors proposed use of dynamic Bayesian network for integrating multi-cue in tracking along with hidden variables. Linear assignment problem algorithm was then applied for model matching. Probability based models such as Hidden Markov model [158–160] has also been used extensively for object detection and tracking in videos. Many researchers have embarked upon use of face clustering and face tracking because they can provide promising information and constraints to each other. In this respect, Wua et al. [159] introduced coupled Hidden Markov Random Field model for modeling face clusters, face tracking and their interaction simultaneously. This has been done by coupling two Hidden Markov random field model. The joint optimization of cluster labels and face tracking was done using constrained clustering and optimal tracking algorithm. Other examples who have used hidden Markov model for object detection in videos includes work of Premaratne et al. [158] and Vojir et al. [160]. In [158], authors have proposed a centroid tracking approach using discrete hidden Markov model for dynamically hand gesture recognition. The proposed system tracks hand signs to evaluate how random signs were recognized by a computer. The system initiates gesture when there is no significant motion of centroid and ends it when centroid comes to a halt; in between actions were considered as single dynamic gestures. The system operated like a dynamic gesture recognition system which uses Hidden Markov model for classification. The authors in [160] introduced a technique for multiple tracking using hidden Markov model. Their technique utilized hidden Markov model for fusion of non-uniform observables and an on-line learned high-precision detector was exploited for pose prediction of multiple complementary trackers. The Hidden Markov model was trained to estimate the state of the tracker in an unsupervised manner.

Syntactic approaches have been defined using domain specific predicates and functions which act as a basis for facts about the knowledge base's domain. Approximate reasoning is one of the important methods of syntactic approaches for object detection and tracking in videos. In [161], a method of action recogonition was presented by Benitez et al. in videos using approximate reasoning and finite state machines. This approach modeled detected objects as a linguistic element of video sequence. Mealy machine was used as an analysis technique for representing predefined actions. Experimental results had proven promising results on real traffic video sequences. The researchers in [162] designed vehicle detection grammar for vehicle detection and handling partial occlusion. The vehicle detection grammar was implemented using constructs of Support vector machine grammar. In the first split, they classified vehicle into constituent parts called semantic parts. While in the second split semantic parts were further divided into sub-parts automatically which was further modeled into grammar model. Then the grammar model was trained using appropriate training procedure to obtain training parameter and thereafter detection of vehicle was done with respect to grammar model.

7.3. Expert system based approaches

Expert system is a powerful tool which mimics decision making ability of humans [163]. The use of expert system for recognition of handwritten digit was address in the work of Alvarez et al. [164]. The method tends to consider numeric handwritten characters decomposed into vertical and horizontal strokes and dynamic zoning was used to extract features. The knowledge base used by the algorithm was developed from a decision tree structure in which all well formatted strings with digits definition was stored. At last, the system used inference engine to achieve digit recognition by matching unknown digits with well-trained knowledge base. The experiments were carried out on MNIST handwritten database and results were encouraging for the proposed approach as compared to conventional approaches. The research work by Yaya et al. [165] presents expert system for detecting broken driving rules for safe driving system. The method uses rule matching algorithm for checking information about the car, driver and environment against the typical driving behaviour, driver conditions etc. The paper has applied rule matching algorithm based on Rete algorithm. The system was developed for deciding whether it was necessary to show driver driving recommendation or not. The driver profile was stored in the knowledge base. The predicted data and information stored in knowledge base was used by expert system to make decision. The decision task consists of two modules namely The “Rule and Data Element Selector” and the “Inference Engine”. The “Rule and Data Element Selector” was used to detect broken driving rules, changes in the current driving behaviour from the typical driving behaviour and uncommon driver conditions, etc. The detection of irregularities was passed to inference engine which decides whether to show driver recommendation or not, which was based on reaction of the driver to an already given recommendation.

7.4. Tracking-by-detection

Object tracking in realistic scenario is often a very challenging task. To improve the performance of object tracking in real time environment, tracking by detection has become popular. Tracking by detection consists of training a classifier online, inspired by statistical machine learning methods, to predict the presence or absence of the target in an image patch. This classifier is then tested on many candidate patches to find the most likely location. Both detectors and trackers can work simultaneously aiding each other to improve long term detection or tracking accuracy. Impressive performance was achieved in proposed work of Kalalal et al. [200]. A long term tracking framework decomposed into three tasks of tracking, learning and detection was proposed. Every task was to be addressed by one component which operated concurrently. The tracker estimated the object from frame to frame. The detector independently scanned all the frames to localize all the observed appearance. The performance of both tracker and detector was evaluated by learner for estimating detector's error and updated those errors to avoid them in near future. Two types of expert i.e. P-expert (recognize missed detections) and N-expert (recognize false alarms) were employed to analyze the response of detector for false alarms. The P-N learning paradigm was incorporated for evaluating detector in every frame of the video. The training set was augmented with the estimated alarms of the detector and re-training of the detector was done to avoid these errors in the future. Motivated by the achievement of tracking by detection approach, Breitenstein et al. [201] employed particle filter framework for multi-object tracking in videos. They implemented first–order Markov model and incorporated the output of detector to be used as an observation model. The algorithm will initially assess the detection in each frame and one of these will be selected for guiding the tracker for a particular target. The classifier was trained for each target and the intermediate, continuous confidence measure of object detector was also exploited. Output of both output was integrated finally by tracker into observation likelihood. More studies employing tracking by detection approach for object tracking in videos are available at [202–204].

7.5. Ensemble tracking

The idea of ensemble methodology is to build a predictive model by integrating multiple models. Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. Shai Avidan [205] proposed an ensemble tracking framework for object tracking by incorporating combination of weak and strong classifier. The methodology works by constantly updating a collection of weak classifiers to separate the foreground object from the background. The weak classifiers can be added or removed at any time to reflect changes in the object appearance or incorporate new information about the background. Each weak classifier was trained on positive and negative examples where examples coming from the object are considered positive and examples coming from the background as negative examples. The pixels of an image were mapped to feature space. 11 dimensional feature vector that is formed by the combination of local orientation histogram and pixel colors was used by classifiers. The ensemble of weak classifiers is combined into a strong classifier using AdaBoost. The strong classifier was then used to label pixels in the next frame as either belonging to the object or the background, giving a confidence map. The peak of map corresponded to object's location and was found using mean shift algorithm. Temporal coherence was maintained by updating the ensemble with new weak classifiers that are trained on-line during tracking. Prior knowledge about the object to be tracked was incorporated the tracker as one or more weak classifiers that participate in the strong classifier, but cannot be removed in the update stage. The work was further improved to handle occulusions by addition of particle filter. The approach of Qinxun Bai et al. [206] differs from the other approaches as instead of trying to compute deterministic optimal weights for the weak classifiers, they characterized their uncertainty by introducing Dirichlet distribution, and draw random samples to form a randomized voting classifier. The parameters of Dirichlet were updated in Bayesian manner. This has helped strong classifier to adapt to the varying appearance, backgrounds, and occlusions. The randomized ensemble tracker was tested in experiments on numerous tracking sequences, that demonstrated robustness of proposed even without motion prediction. A summarization of various ensemble based classifier methods was first presented by Lior Rokach [207]. The paper discusses the ensemble process in detail and conventional ensemble based methods applicable in statistics and AI. Studies on methods for combining base classifiers outputs as weighting methods and meta-learning methods are discussed. Weighting methods are useful if the base-classifiers perform the same task and have comparable success. Meta-learning methods are best suited for cases in which certain classifiers consistently correctly classify, or consistently misclassify, certain instances. The survey further discusses studies that differ on diversity generation, provides an overview of studies on selection of how many and which base classifiers to use. The survey also highlights studies on multi-strategy ensemble tracking. [208] is one of the latest attempts to summarize the growth of ensemble based classification methods. The study has classified ensemble methods into conventional ensemble methods such as bagging, boosting and random forest, decomposition methods, negative correlation learning methods, multi-objective optimization based ensemble methods, fuzzy ensemble methods, multiple kernel learning ensemble methods and deep learning based ensemble methods. The discussion of deep learning based ensemble methods is one of the highlight of the manuscript.

Input: Image patches, historical data, Random forest parameters, Dirichlet –multinomial parameters
Output: Target position, current labels, new random forest parameters, new Dirichlet-multinomial distribution parameters.
For Iteration =1
Step 1: Initialize (new random forest parameters)=Given Random forest parameters.
while ((no improvement in solution) and (iteration< predefined threshold) do
Step 2: Perform classification on the current frame based on (random forest parameters) for current iteration to generate new label set.
Step 3: New target location was estimated based on classification results.
Step 4: The foreground grid was determined once the location is estimated.
Step 5: The labels were accepted based on Dirichlet –multinomial parameters.
Step 6: Random forest is retained
end while
Step 7: update new distribution parameters
Step 8: update Dirichlet –multinomial parameters.

C8:-Pseudo Code for Robust Discriminative Tracking via Structured Prior Regularization [218]

In order to overcome inaccuracies during tracking process Yuanhao Yu et al. [218] proposed another ensemble based technique by forming a reliable generative model to support learning process. The proposed algorithm consists of three modules, the appearance model, the structure model, and the optimization algorithm. The appearance model was responsible for preserving similarity in appearance by maintaining a built-in semi-supervised Random Forest classifier with Dirichlet-multinomial parameters having loss function L_a. The structure model was used to localize the target by identifying the most similar structure pattern using its loss function L_s. Tracking was formulated as a prior regularized semi-supervised learning problem, in which objective function was defined as L_s+ L_a. A Global optimization function was obtained by making use of the structure loss function as a regularization term added to the appearance loss function. Optimization made use of a decision maker (DM) to find the Pareto optimal solution for the proposed objective function.

7.6. Correlation filters based studies

Correlation filter based trackers model the appearance of objects using filters trained on example images. The target is initially selected based on a small tracking window centered on the object in the first frame. Tracking and filter training work together thereafter. The target is tracked by correlating the filter over a search window in next frame; the location corresponding to the maximum value in the correlation output indicates the new position of the target. An online update is then performed based on that new location. Correlation filters can track complex objects through rotations, occlusions and other distractions at much faster rate than state-of the-art techniques. Naive methods for creating filters, such as cropping a template from an image, produce strong peaks for the target but also falsely respond to background. Average of Synthetic Exact Filters (ASEF), Unconstrained Minimum Average Correlation Energy (UMACE), and Minimum Output Sum of Squared Error (MOSSE) [209,210] produce filters that are more robust to appearance changes and are better at discriminating between targets and background. In [209], the targets appearance was modeled by adaptive correlation filters, and tracking was performed via convolution to create a fast tracker in which correlation is computed in the Fourier domain using Fast Fourier Transform (FFT). Fourier based techniques are extremely limiting therefore more powerful features like kernel methods for correlation filters have been proposed in [211,212]. Kernel filters tend to ignore the spatial structure of image. J. F. Henriques et al. [213] proposed a new variant of the Kernelized Correlation Filter (KCF) tracker to cope with changes in natural images. This classifier was typically trained with translated and scaled sample patches based on Histogram of Oriented Gradients (HOG) features instead of raw pixels. They also proposed a fast multi-channel extension of linear correlation filters knows as Dual Correlation Filter (DCF) with low computational complexity.

8. Datasets

Due to numerous affordable commercial sensors released, object detection and tracking in videos have gained more and more popularity. In recent decade a large number of datasets have been designed by researchers to evaluate performance of different algorithms. For this purpose, several datasets are made publically available to researchers and developers. The recent availability of object detection and tracking datasets has led to many new discoveries in computer vision field. Selection and availability of benchmark dataset is first major step towards global acceptance of a specific algorithm. For benefit of research community, authors decided to gather a list of available dataset for evaluation of algorithms in domain of soft computing. [214] is a multiple target tracking dataset that can be used for evaluation of various types of applications such as pedestrian detection, short term tracking etc. It consists of existing public videos, latest challenging sequences, and is open for new submissions. This framework provides various challenges with subsets of data for specific tasks such as sport analysis, 3D tracking etc. It also provides evaluation tool for different measures such as recall, precision provided in Jason format. The encoding color information benchmark [176] for visual tracking compiled 128 video sequences with ground truth and challenge factor annotations. It also encoded 10 chromatic models with 16 selected state-of-the-art visual trackers. This purpose of developing this dataset was to guide and motivate young group of researchers for research on encoding color in visual tracking. Various challenging factors that affect visual tracking performance were also included. NUS-PRO dataset [215] stands for NUS- People and Rigid dataset designed for usage in object tracking algorithm evaluation. It comprises of 365 pedestrian and rigid object image sequences. This dataset is publically available for research on tasks such as object detection, object classification, etc and contains sequences captured via moving camera.

The Visual Object Tracking (VOT) challenges [216] provide visual tracking community with a precisely defined and repeatable way of comparing short-term trackers as well as a common platform for discussing the evaluation and advancements made in the field of visual tracking. The goal of various challenges is to build up a repository of considerable benchmarks to push forward research in visual tracking. VOT benchmarks datasets are available in different variants. VOT 2013 comprises of 16 video sequences which were chosen based on clustering visual features of object and background. It contains image directories for each sequence and annotation of the axis aligned bounding box marking the object per frame. VOT 2014 comprises of 25 video sequence from VOT's 2013 and ALOV dataset. The annotation of the sequence was done using

Table 6
Commonly public available object detection and tracking datasets.

Dataset	Format	Year	Type of problems	Purpose & Description	Source of video	Description of Ground Truths	Areas of applications	All studies
Wallflower [166]	Images	1999	Several different background maintenance algorithms are tested on seven different image sequences.	This dataset contains 7 video sequence and one ground truth for each video sequence which represents problematic scenario for background maintenance	Recorded indoor and outdoor videos	Hand-segmented ground truth	Dynamic Object Detection in videos with Real Scenarios [30], object detection in videos [62]	[30,32,56,62]
CAVIAR [167]	Videos	2002	City Center Surveillance and automatic Analysis of Customer Behavior	Designed for development of algorithms to describe and understand video scenes.	Recorded indoor and outdoor videos	Bounding Boxes and action annotation	Moving object detection and tracking in videos [92,116], tracking multiple targets in videos [103], Human behavior analysis from videos [126].	[40,92,103,126,116]
AVSS [168]	Videos	2007	Developed for detecting abandoned objects.	This is a dataset for event detection and is a sub-set of the i-Lids dataset	Videos from CCTV footage	object and action labels, in XML	Object detection based on behavior patterns [72]	[72]
Weizmann human action dataset [169]	Videos	2007	It is the first human action video datasets	Formed by a unique long sequence of around 6000 frames, displaying different people, wearing different clothes, and performing four activities: running in place, waving, running, and walking.	Recorded indoor and outdoor scenes	Temporal annotations (name of the actions for each frame)	detection of human actions [75]	[75]
VidTIMIT [170]	Videos	2009	Video and corresponding audio recordings of 43 people, reciting short sentences.	This dataset is proposed for automatic lip reading, multi-view face recognition, multi-modal speech recognition and person identification.	Videos from different sources	Simple action annotation	Object tracking in videos [113]	[113]
UCF50 [171]	Videos	2010	UCF50 is an extension of YouTube action dataset providing more action example	CF50 is an action recognition dataset with 50 action categories, consisting of realistic videos taken from YouTube in 2010.	Videos from web (indoor and outdoor)	Frame ranges, bounding boxes, action annotation, in VIPER	Human action recognition [123,124]	[123,124]
CVPR2013 [172]	Videos	2013	Provides realistic challenge to tracking in videos	CVPR2013 tracking benchmark which contains 50 fully annotated image sequences.	Videos from different sources	Bounding boxes and action annotation	Object tracking in videos [50–52]	[50–52]
CVPR2015 [213]		2015		CVPR2015 tracking benchmark contained ground truth object positions and sequence of attributes were introduced for performance analysis. integrated most of publically available trackers into one library with uniform inputs and outputs				

Table 6 (*Continued*)

Dataset	Format	Year	Type of problems	Purpose & Description	Source of video	Description of Ground Truths	Areas of applications	All studies
ETH [173]	Videos	2011	Contains three video sequence and variation lies in pedestrian appearance.	Commonly used by researches for human detection	Recorded indoor and outdoor videos	Bounding Boxes and action annotation	Multi target detection for tracking applications [48]	[48]
ETISEO [174]	Videos	2011	Created for improvement of robustness in video surveillance algorithm and dependencies between algorithms and conditions of use.	Videos are grouped into five topics: apron, building corridor, metro, road and building entrance.	Recorded indoor and outdoor videos	General attributes (sequence id, etc.), scene attributes (In/Outdoors sequence), video acquisition (camera calibration, etc.),	Object tracking in videos	[47,67]
RGB-D People [175]	Videos	2011	Dataset contains RGB-D frames acquired in university hall from three vertically mounted Kinect sensors	Provides on-line evaluation of both RGB and RGBD tracking algorithms for generic objects. 100 RGBD tracking datasets along with RGBD tracking software with source code. An on-line submission script that allow to evaluate tracking algorithm in their framework.	Recorded indoor and outdoor videos	Per-frame bounding box annotations of individual people, together with visibility measure provided in .jason format.	Tracking people in RGBD videos [49]	[49]
OTCBVS [177]	Videos	2011	Used for testing and evaluating novel and state-of-the-art algorithms	The benchmark contains videos and images recorded in and beyond the visible spectrum.	Infrared and Thermal videos	Bounding boxes	Object tracking in videos [106]	[106]
BMC [178]	Videos	2012	Developed for detection of moving objects, motion estimation.	Synthetic videos and real videos for evaluation purposes with encrypted ground truth.	Recorded indoor and outdoor videos	Frame ranges with action annotation	Object detection in videos [32], automatic motion analysis for dynamic object detection in videos. [121]	[30,32,33,121]
Changedetection.net 2014 [179]	Videos	2014	2012 problems and incorporated additional challenges: Challenging Weather, Low Frame-Rate, Night, Air Turbulence.	Used for testing and ranking algorithms for change and motion detection. Matlab and Python programs are available to compute metrics described on 2012 and 2014 results.	Recorded indoor and outdoor videos	Bounding boxes and ground-truth contain 5 labels: static, hard shadow, outside region of interest, unknown motion, motion	Object detection in videos with scene-specific data [41]	[41,32]
PETS [180,181]	Videos	2001–2 2007, 2015	This dataset is used for testing visual surveillance algorithms.	This dataset can be used for detecting abandoned objects.	Recorded indoor and outdoor videos	Bounding boxes	Object detection for surveillance applications [116], object-tracking in videos [112], object detection in videos [61,62].	[46,56,61,62,103,112,116]

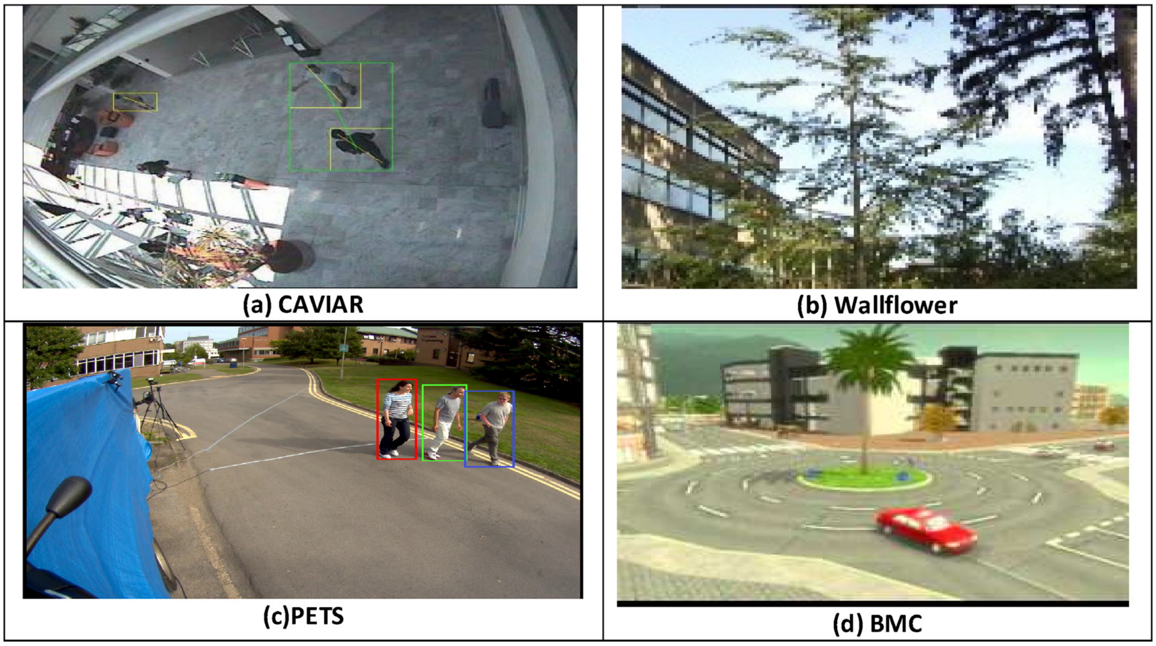

Fig. 9. Sample images from available Datasets.

rotated bounding boxes. VOT 2015 comprises of 60 video sequences from OTB2 dataset, ALOV dataset, non-tracking dataset, computer vision online, data wrangling, etc. VOT 2016 video sequence was very similar to VOT2015 but VOT2016 ground truths were more accurate than VOT2015. Some of the available benchmark datasets that can act as a brief guideline for researchers for evaluation and performance comparison of their algorithms are presented in this section and Fig. 9 shows some sample images captured from available datasets. Almost all datasets are shipped with truth values in form of binary frames and all of datasets are publically available except a few where user is supposed to sign a contract. The datasets also provide facility for downloading source codes for some of the implemented methods which can be very helpful for researchers in building their own solutions. A collection of all datasets used by researchers for evaluation of various soft computing based solutions is presented in Table 6. The table also lists year in which dataset was made available along with specific type of problem for which datasets were designed and tested. The table also provides some basic description of dataset and source from where videos are captured. Elaborations on available ground truth and types of applications on which these datasets have been tested are also listed. Information about the specific areas of application of dataset and details of study on which it has been used is provided in column 8. The last column provides reference of studies on which specific dataset have been used.

In summary, Wallflower and PETS provides videos dedicated for video surveillance. CAVAIR dataset is dedicated to automatic analysis of customer behavior and city center surveillance. BMC dataset provides synthetic and real videos and in addition videos are separated into two folders for learning and testing. Changedetection.net contains infrared videos along with scripts for testing and evaluating own methods. OTCBVS dataset also contains videos and images recorded in and beyond the visible spectrum while AVSS contains CCTV footage. Although, in older dataset ground truth provided were limited to manual annotations but the majority of modern datasets provide high quality ground truth images or frames along with other details like bounding box parameters, regions of interest, motion etc. It can be observed that PETS, Wallflower CAVAIR along with BMC dataset are more commonly used datasets by researchers as compared to others.

9. Challenges in object detection and tracking in videos

Accurately detecting and tracking object from a video sequence is a challenging task because of the fact the object can have complicated structure and can change shape, size or orientation over subsequent video frames. Designing efficient and accurate system is always a big challenge. Fig. 10 shows various issues and major challenges for moving object detection process in videos. These challenges arise because of occlusions, short comings of capturing devices, variations in scenes or appearance, shadows of the objects appearing in the frame. These tend to degrade the performance of the developed algorithm and results of detections may be poor. To overcome these challenges, the developed algorithm must take care of these issues while proposing solutions for specific applications. The major design issues are summarized in this section that can act as various challenges for researchers to focus upon.

1. **Low quality of Image** capturing tool results in generation of noisy or blurry image that in turn leads to false detection. The image could also be noisy because of weather conditions like rain, fog etc. The system shall be able to work in noisy images and able to detect the objects in videos with precise boundaries. The quality of camera deployed for capturing images need to be considered together with the weather conditions.
2. Camera jitter makes the captured object look blurred with prolonged boundaries. The proposed system shall provide effective ways to handle **camera jitter** that occurs due to high velocity winds blowing at time of image capturing. The detection method shall be able to overcome this limitation.
3. **Video captured on moving cameras** like cameras installed on top of vehicles add another dimension to this already challenging area. The movement of camera needs to be simulated by the algorithm for effective and accurate object detections. The problem of moving object detection within moving camera is one of the most happening areas being explored by researchers.
4. Changes in illumination can occur because of presence or disappearance of a light source in background for eg bulb, tube light, sun etc Rapid **Illumination changes** in the scene of interest lead to false detections in consecutive frames or over multiple frames.

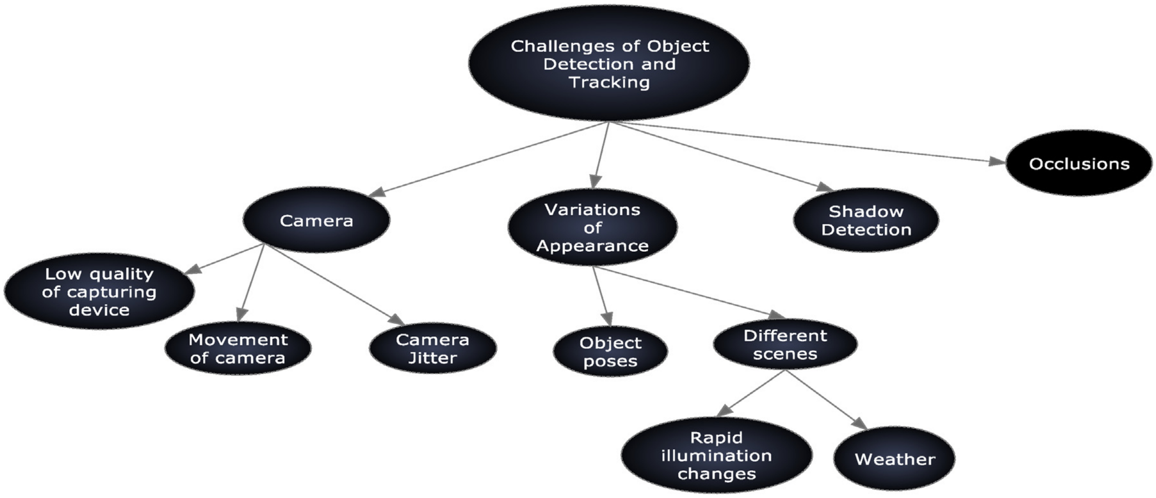

Fig. 10. Major issues in object detections and tracking.

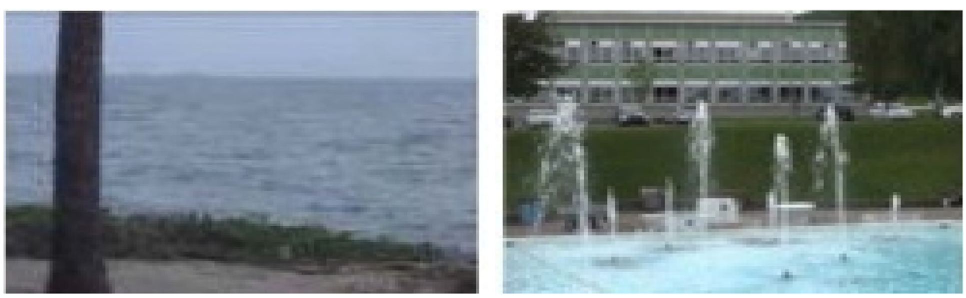

Fig. 11. Example of dynamic background (a) moving water in sea (b) fountains.

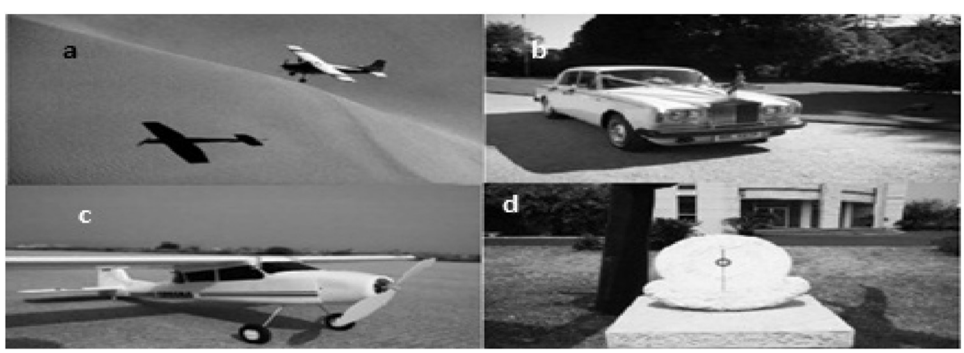

Fig. 12. Sample images showing shadows (a) Flying plane with its shadow (b)moving car (c)shadow of standing plane (d) shadow of object (top left to bottom right).

The developed solution shall be able to work on different levels of illuminations.

5. Rapidly varying background also complicates detection process. The dynamism in the video may result because of minor movements in background because of heavy winds blowing; object moving quickly, underwater images etc. see Fig. 11. Varying background makes false detections more prevalent in the final detected object frame. The designed system shall be able to simulate effects of **dynamic backgrounds.**
6. Shadows coming from background object or moving object can be detected as foreground, Fig. 12. Shadows must be clearly detected separated from correct object. Some suitable method must be employed for this purpose, there are studies to help the researchers overcome this issue and readers are advised to go to [182,183] for further details.
7. **Occlusion**: In a video sequence object of interest may fall behind another object which is present in the current scene. This is known as occlusion. Fig. 13 shows some examples of occluded objects. Occlusions may occur in partial or full and in both cases can be treated as a major research challenge. Occlusion result in overlapping of detected boundaries and wrong calculation of number of objects. Some methods have been applied successively to deal with occluded objects but it remains still a challenging task to overcome this issue and make appropriate detections.

10. Analysis and inferences

This section provides some insights about the progress of various techniques for benefit of audiences and some analysis done by

Fig. 13. Examples of Occlusions (a)persons occluding each other (b)person completely occluded by another person, (c)-(e) parts of human body occluded partially or fully (top left to bottom right).

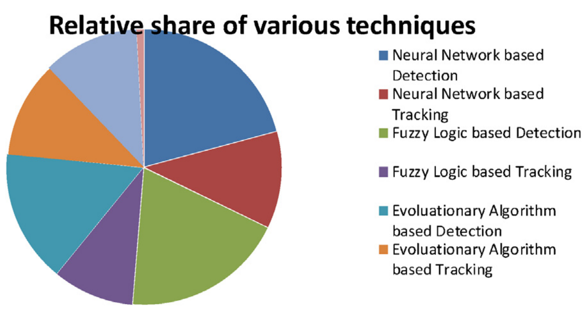

Fig. 14. Relative share of various approaches.

authors is discussed. The first analysis portrays relative share of a specific technique among other techniques. It is presented to make audiences aware about growth of different techniques with respect to each other in the present scope in terms of share in publications. Fig. 14 gives an observation that neural network based object detection and tracking approaches have been most targeted area as compared to other techniques. This comes from strength of neural network in their ability to provide generalized solutions through an architecture which is able to learn to improve its performance. Techniques based on fuzzy logic and evolutionary approaches also enjoy a fair share however it is presented that hybrid models are least commonly used models for object detection and tracking in videos, therefore, research efforts focused on development of applications using hybrid techniques to improve object detection and tracking performance can be expected in future. This can also be observed from the chart that overall number of studies on object detection methods is more as compared to object tracking methods and this trend follows among individual subcategories also. Under hybrid techniques, development of object detection methods has been more rapid as compared to tracking methods. This may be because of computational complexities associated with implementation of these studies. This is an indication of the fact that more number of studies for tracking are likely to be seen in the near future targeting hybrid techniques.

Fig. 15 presents the statistical summary of the different datasets that have been used for evaluating object detection and tracking techniques in videos. Various soft computing techniques are represented along X-axis while their relative contribution in percentage is projected along Y-axis and legend entries correspond to specific datasets. Those datasets whose use is above a specific threshold value have been included for brevity and to make results more meaningful. The observation shows that majority of evaluations have been done by researchers using videos captured by themselves or from videos available from the web. This is true across all the four categories. PETS with its different variant is most commonly used datasets as compared to others for almost all techniques while use of UCF50, RGB-D people dataset and real time videos are least explored by researchers. Use of CAVAIR dataset is common in various application developed with evolutionary and hybrid approaches. Neural network based techniques have been tested across almost all datasets while studies based on fuzzy logic, hybrid or evolutionary algorithms have been confined to limited number of datasets. This can be concluded that more testing of applications developed in field of hybrid or evolutionary algorithms is required with diverse datasets which can be another research area for future exploration.

Fig. 16 depicts correlation between soft computing techniques and applications on which they are employed in terms of number of papers published. The graph is useful in demonstrating strengths of various approaches in proposing solutions for specific application. Different applications have been categorized into four categories for brevity and to make the graph meaningful and concise. Type A consists of surveillance and traffic based applications like video, traffic, marine flow monitoring, vehicle stopped in parking etc. Type B consists of applications developed for pedestrian, human motion detection and tracking; type C groups together application developed for detection and tracking of any other objects like animal, tumor detection etc; Type D consists of applications developed for real time videos or sports sequences etc. According to graph, neural network and fuzzy logic based techniques are more common solutions for almost all types of applications, this may be because of their relative simplicity and the fact that these theories have relatively longer history and dissemination. On the other hand, progress of evolutionary approaches has been less common overall apart from application type C where they have shown remarkable promise. The contribution of hybrid techniques shows rare utility for various applications except for applications of type C and may be seen as future area for active research. The graph also suggests that development of solutions for real time applications can be another future area of research as number of studies in this field for all categories is almost negligible.

Another form of analysis, in which various studies were compared for their skills to handle specific research challenge are presented in Table 7. Table 7 provides analysis of studies w.r.t various challenges along with various performance analysis metrics pursued on object detection and tracking in videos. The table has grouped various studies arranged in chronological order along with major techniques employed. The first column of table describes study employed along with major technique used in the study for handling specific challenges. Various challenges have been tabu-

Table 7
Summary of the challenges along with evaluation metrics used for soft computing based strategies for object detection and tracking in videos.

	Challenges undertaken	Dataset	Performance Analysis*										
		Qualitative	Quantitative			Other parameters							
Study with employed Strategy	Camera Jitter		Noisy Image	Dynamic Background	Illumination changes	Shadow detection	Occlusion		Others	F	P	R	
[92] Tracking based on fuzzy membership function, 2006	×	√	×	√	×	×	Background clutter	infra-red image sequences and CCD images	√	×	×	×	RMSE = 4.062 Normalized error rate = 0.069
[33], Self organizing Map, 2008	√	×	√	√	×	×	camouflage bootstrapping	five different sequences for video surveillance	√	0.95	0.89	0.84	Similarity=0.80
[91], fuzzy hostility, 2009	×	√	√	√	×	√	–	time-ordered frames of video sequence of F-16 fighter Plane	√	×	×	×	Computational time = 3.33ms
[46], Convolution neural network, 2010	×	×	×	√	×	√	Scale and View Changes	challenging video sequences	×	×	×	×	Position error=2.88 (occlusions) Position error = 7.85 (indoor)
[127], Neuro-fuzzy approach, 2010	×	×	√	√	×	×	×	publicly available sequences	√	0.9	0.9	0.87	Computational complexity=$O(n^2NM)$ n is the number of weight vectors used to model each pixel and N × M is the image dimension.
[60], Fuzzy Particle filter, 2011	×	×	√	√	×	×	Background clutter	indoor and outdoor sequences taken by fixed camera	×	×	0.6886	o.8843	Number of different fuzzy measures like fuzzy divergence, fuzzy minmax, contrast de-enhancement, fuzzy NAD, GTI etc.
[49] radial basis function network, 2011	×	×	√	√	×	√	dynamic change in object appearance, scale/size	fixed camera sequences	√	×	×	×	RMSE = 2.1 Average deviation in terms of Center = 2.4; Width=2.9; Height=1.8
[98], Simulated annealing based optimization, 2012	×	√	√	√	×	×	–	CAVIAR (2010)	×	0.7075	0.6209	0.8223	AFR 0.7223 AOC = 0.8484
[87], fuzzy logic, 2012	×	√	×	√	×	√	Dynamic change in Objects	real life scenarios	√	×	×	×	RMSE position=8.85px RMSE rectangle size = 4.88 px Processing time per cycle = 22.64 ms
[114] particle swarm optimization,2013	×	×	√	√	×	√	–	CAVAIR, Oxford and PETS 2009	√	×	×	×	Oxford | PETs MOTA|84.8% |83.0% MOTP|80.5%|86.4% MT |79.6% |94.7% ML |18.7 | 5.3% PT |1.7% | 0.0%
[86]. Fuzzy particle filter, 2013	√	√	√	√	√	√	Background clutter	Video sequences obtained from in vehicle camera	√	×	×	×	average time=0.142ms MSE=4.1
[77], Fuzzy rule-based reasoning, 2013	×	√	√	×	×	×	–	Soccer video dataset	√	78.67	76.12	81.78	No. of detectable events = 7
[123],multi-feature PSO,2014	×	√	×	√	×	√	Complex Object shape/motion Background clutter	PETS, FLIR dataset.	√	×	×	×	No. of failures = 39/1505 Accuracy=8.334; TRDR, OSR etc.
[32], Self organizing Map, 2014	×	√	√	√	×	×	Camouflage	Wallflowers, PETS	√	0.8020	0.8005	0.8138	PSNR = 42.6204 Dscore = 0.0073 Ssim = 0.9638
[132], Neuro-fuzzy approach, 2015	×	√	√	√	√	×	Camouflage, dynamic objects,	BMC	√	0.7850	0.8232	0.7543	PSNR = 42.4361 Dscore = 0.0060 Ssim = 0.9573
[45], mean shift and radial basis function neural networks 2015	×	×	√	×	×	√	Scale variations, background clutters	PETS 2007	×	×	×	×	FET = 450 NNTT = 2.43 s
[34], Self organizing Map, 2016	√	×	√	√	√	×	stationary objects of interest became dynamic	BMC	√	0.7419	0.6533	0.8582	Psnr = 36.8502 Dscore = 0.0054 Ssim = 0.9559
[54],deep Learning, 2016 (NIN)	×	√	√	√	×	√	Background clutter	CVPR2013	√	0.611	×	×	OPE=0.452
[53],deep convolutional neural network, 2016	×	×	√	√	×	√	Background clutter	CVPR2013	√	0.684	×	×	OPE=0.472
[120], adaptive discrete swarm optimization, 2016	√	×	√	√	×	×	changing scale of object, unexpected object motion	video from Visual Tracker Benchmark	√	×	×	×	error position (pixel) per frame = 78 0.3242 Processing time per frame = 2 0.870ms
[43], deep convolutional neural network,2016	√	×	√	√	√	×	Night videos, Bad weather Low frame rate	2014 CDnet	√	0.9046	×	×	$F_{Baseline}$ = 0.9813 F_{Jitter} = 0.9020 $F_{DynamicBG}$ = 0.8845 $F_{Shadows}$ = 0.9454 $F_{BadWeather}$ = 0.9264 F_{Night} = 0.7565

F → F-measure MT → Mostly tracked AFR → Average Fragmentation (AFR) FET → feature extraction time AOC → Accuracy of Object Count P → Precision PT → Partially tracked MSE → Mean signal to noise ratio TRDR → Tracker detection rate PSNR → Peak signal to noise Ratio R → Recal MOTA → multiple object tracking accuracy OSR → Occlusion success rate NNTT → Neural network test time ML → Mostly lost MOTP: Multiple object tracking precision OPE → Overlap precision SSIM → structural similarity
*Values as extracted from result section of original papers

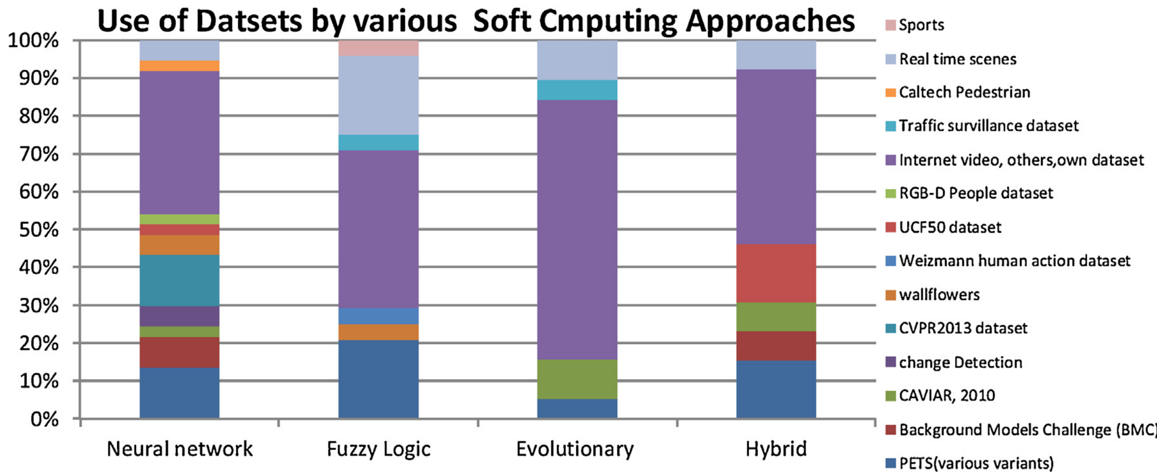

Fig. 15. Types of datasets used in various soft computing based approaches for object detection and tracking in videos.

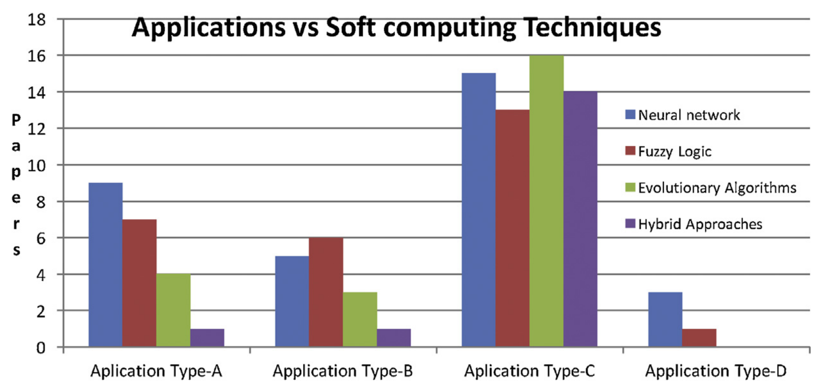

Fig. 16. shows frequency of publication (along y-axis) under each category of soft computing approaches (color coded) used for different categories of application (along x-axis).

lated column wise and intersections demonstrate applicability of the study to a particular challenge. In order to provide a deeper perspective into effectiveness of a particular approach, values of various parameters used for evaluation in their result section are also listed. Good value of a particular metric indicates that proposed publication has been successful in handling listed challenges. It can be observed that a large number of metrics are used for evaluation of diverse soft computing techniques. Precision, Recall, F-measure are most common metrics for both detection and tracking where as object tracking methods have been commonly evaluated using RMSE, computational time, no. of failures, etc. Although these metrics are commonly practiced but measures based on bounding box comparisons also find a good number of references in studies. The indicative values of these parameters are also specified as part of table along with their names. Readers are advised to refer to individual studies for further details and discussions. It can also be observed that a considerable amount of object detection and tracking related publications have been evaluated qualitatively. It can be observed that illumination change along with handling of dynamic background is widely targeted research challenge and finds implementation with all major soft computing approaches. Handling of noise and occlusions in videos is the second sought after research challenge for which major studies based on fuzzy logic and evolutionary based algorithms have been used. There has been little progress in handling of camera jitter and shadow detection although some studies have tried to provide solutions for these areas but it still remains a major research challenge. The purpose of this classification is to provide researchers with a quick guide to make them aware as what sorts of algorithms have already been implemented to overcome certain challenges and what algorithms to explore for further research. Furthermore, when real time performance is desired, object detection and tracking in video sequences is still an open research field with various issues/challenging research opportunities. There have been substantial improvements in results with advent of techniques like deep learning, deep convolution networks and hybrid approaches but this topic still remains a major research challenge.

10.1. Selection of algorithms: some guidelines

Selection of algorithm for a specific application is tough task and to increase effectiveness of object detection and tracking process in videos, existing soft computing based approaches should be selected carefully based on purpose and characteristics of the applications. Some guidelines are discussed here in this section.

It can be inferred from analysis done that for applications where critical situations such as sudden illumination changes, random noise can affect detection process, feed forward neural network based approach is used as an efficient tool for classification example in [22,25]. For applications that suffer from dynamic background, annoying noise etc. self organizing maps have been successfully employed, corresponding studies can be found at [28,29,32–35]. Self organizing map approach is capable to detect anomalous objects in videos but major limitation of self organizing map is that they are computationally much expensive. Conventional Neural Network are not much suitable for object detection in real time applications because of long periods of computation in the training process as is clear from articles [24,26,27]. These studies have used competitive Neural Network approach to perform parallelized object segmentation to achieve better object detection in complex real time video sequence. Another direction towards efficient object detection and tracking to handle background clutter, occlusions, dynamic background or noisy images were given by deep learning neural network at studies [50–53]. Deep learning based convolutional neural networks can also be used to handle night videos, bad weather, low frame rate etc. Fuzzy logic had been great success in the past for applications that suffer from high level of uncertainties. Therefore in [68], fuzzy integrals were applied in aggregation step to provide better classification of pixels either to foreground or background. Type-2 fuzzy Gaussian mixture model approach offered capability to handle higher dimensions of uncertainty which resulted in smoother classification results in articles [57,59,60]. Fuzzy particle filters implemented by number of studies in [80–83], it increases robustness of the system as it can handle system complexity by compromising low number of particles. Fuzzy classifiers can also be safely used for detection and tracking in videos under several dynamic environment conditions.

Optimization approaches for object detection and tracking objects in videos have been proposed aiming to improve classification of pixels either to foreground or background. Entropy based simulated annealed background subtraction method in [98] is used for optimization because of its effective exploitation within search space and capabilities to avoid local minima. Optimization studies like Particle Swarm Optimization have been implemented for tracking multiple objects. The success rate of these applications has been found to be better than Kalman filter or other contemporary techniques. Multi feature particle Swarm Optimization approach offers a solution towards low accuracy of conventional Particle Swarm Optimization approach and can mitigate effects of complex object shape, object clutter, motion etc. as shown in [123,125]. The major advantage of method is based on strong global optimization ability, which makes it suitable for real time processing and detect objects of any shape or size. Genetic Algorithms could be extended to model background effectively to overcome illumination changes using edges and intensity values of frame as shown in [99,100]. Hybrid methods of object detection and tracking approaches in videos are motivated by the fact because they focus on combining the best properties of two or more object detection and tracking techniques. Recent initiatives such as [126–129,132] etc. shows that researchers take interest in applying hybrid techniques for object detection and tracking in videos to achieve accurate results. They have been successful in addressing issue related to uncertainty in detection process caused by the cited background maintenance phase of Background subtraction approach. The experimental results on real color video sequence showed promising results and redirected researcher's attention towards hybrid approaches of object detection in videos. Hybrid techniques have been very successful in proposing effective solutions in complex field of tumor detection [141–143]. Study [131] based on neuro-fuzzy approach has shown some promise in overcoming challenges posed by shadows, camouflaging.

11. Conclusions and summarizations

In the last decade, the problem of object detection and tracking in videos has received considerable attention of researchers around globe. In the present study, various soft computing based approaches for moving object detection and tracking in videos were presented. Although a comprehensive review of all various soft computing based approaches for moving object detection and tracking in videos is beyond scope of any paper but articles particularly relevant to the topic of study are presented. Research articles utilizing theories of fuzzy logic, neural networks and evolutionary based algorithms have been explored and categorized accordingly. The implication of this study is to provide better understanding of soft computing based approaches for moving object detection and tracking in videos. It is prudent to say that literature presented in this survey provides categorization of various recent articles according to soft computing based approaches which will help researchers to choose an appropriate approach for particular application. Secondly, it presents an insight into failure as well as success of the existing soft computing based approaches which will help budding researchers to have a wider observation on the entire field. Availability of datasets, research trends, issues and future directions are also highlighted. Moreover, based on the characteristics of the application, a brief guideline is provided for the selection of appropriate soft computing based approach. A number of issues to guide future directions of research were also provided in the paper. One of the major challenge with all soft computing approaches is the time taken to converge to a global optimum solution; no such study is available in which comparisons with respect to settling time have been made with two or more soft computing based algorithms. Some new soft computing techniques that shows quick convergence time as compared to techniques like genetic algorithm, PSO etc. like Ions Motion Algorithm [184], Water Cycle Algorithm [185], Water wave Optimization [186], Mine Blast Optimization [187] can be explored. Solutions based on such techniques can be explored for real time video analysis and processing. Preprocessing of videos is also one of the emergent areas that can improve the detection results and a very small number of studies using soft computing techniques have been proposed for preprocessing or enhancement of videos. Studies based on nature inspired algorithms available at [188,189] could be explored for such research directions. The significance of survey may be gauged from industrial, academic and end user level; Academician and young students will be benefitted as image processing and computer vision together with machine learning or soft computing is being studied by large number of students at both graduate and post graduate levels, the concepts presented may inculcate novel ideas in young minds which may be hatched at an early stage. Industry and end user may be directly befitted as more and more explorations in the field will attract researchers to devise better and relatively low cost solutions that are reachable to all sections of society.

References

[1] Q. Ling, J. Yan, F. Li, Y. Zhang, A background modeling and foreground segmentation approach based on the feedback of moving objects in traffic surveillance systems, Neurocomputing 133 (8) (2014) 32–45.
[2] Y. Tian, A. Senior, M. Lu Robust, Efficient foreground analysis in complex surveillance videos, Mach. Vis. Appl. 23 (5) (2012) 967–998.
[3] R. Arroyo, J.J. Yebes, L.M. Bergasa, I.G. Daza, J. Almazán, Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls, Expert Syst. Appl. 42 (21) (2015) 7991–8005.
[4] D. Makris, T. Ellis, Path detection in video surveillance, Image Vision Comput. 20 (12) (2002) 895–903.
[5] M. Shakeri, H. Zhang, Real-Time bird detection based on background subtraction, World Congress on Intelligent Control and Automation at Bejing, China (2012) 4507–4510.
[6] J. Heikkila, O. Silven, A real-Time system for monitoring of cyclists and pedestrians, 2nd IEEE Workshop on Visual Surveillance at Fort Collins, USA (1999) 74–81.
[7] M. Piccardi, Background subtraction techniques: a review, IEEE International Conference on Systems, Man and Cybernetics (2004).
[8] A. Yilmaz, O. Javed, M. Shah, Object tracking: a survey, ACM Comput. Surv. 38 (4) (2006) 13.
[9] T. Bouwmans, F.E. Baf, B. Vachon, Background modeling using mixture of Gaussians for foreground detection: a survey, Recent Patents Comput. Sci. 1 (3) (2008) 219–237.
[10] B. Karasulu, Review and evaluation of well-known methods for moving object detection and tracking In videos, J. Aeronaut. Space Technol. 4 (4) (2010) 11–22.

[11] T. Bouwmans, Recent advanced statistical background modeling for foreground detection: a systematic survey, Recent Patents Comput. Sci. 4 (3) (2011) 147–171.
[12] J.J. Athanesious, P. Suresh, Systematic survey on object tracking methods in video, Int. J. Adv. Res. Comput. Eng. Technol. 1 (8) (2012) 242–247.
[13] C.L. Devasena, R. Revathí, M. Hemalatha, Video surveillance systems – a survey, Int. J. Comput. Sci. Issues 8 (4) (2011) 635–642.
[14] K.A. Joshi, D.G. Thakore, A survey on moving object detection and tracking in video surveillance system, Int. J. Soft Comput. Eng. 2 (3) (2012) 44–48.
[15] J. Raiyn, Detection of objects in motion—a survey of video surveillance, Adv. Internet Things 3 (4) (2013) 73–78.
[16] A.K. Chauhan, D. Kumar, Study of moving object detection and tracking for video surveillance, Inte. J. Adv. Res. Comput. Sci. Softw. Eng. 3 (4) (2013) 911–913.
[17] D. Gowsikhaa, S. Abirami, R. Baskaran, Automated human behavior analysis from surveillance videos: a survey, Artif. Intell. Rev. 42 (4) (2014) 747–765, Springer Netherlands, ISSN: 1573–7462.
[18] A. Sobral, A. Vacavant, A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos, Comput. Vision Image Understanding 122 (6) (2014) 4–21.
[19] S. Jeevaa, M. Sivabalakrishnan, Survey on background modeling and foreground detection for real time video surveillance, Procedia Comput. Sci. 50 (2015) 566–571.
[20] C. Hong Lim, E. Vats, C. Chan, Fuzzy human motion analysis: a review, Pattern Recogn. 48 (2015) 1773–1796.
[21] D.T. Nguyena, W. Lia, P.O. Ogunbona, Human detection from images and videos: a survey, Pattern Recogn. 51 (2016) 148–175.
[22] M.A.A. Dewan, M.J. Hossain, O. Chae, Moving Object Detection and Classification Using Neural Network, Springer, Berlin, Heidelberg, 2008, pp. 152–161.
[23] S. Taleghani, S. Aslani, S. Shiry, Robust Moving Object Detection from a Moving Video Camera Using Neural Network and Kalman Filter, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 638–648.
[24] R.M. Luque, E. Domínguez, E.J. Palomo, J. Muñoz, A Neural Network Approach for Video Object Segmentation in Traffic Surveillance, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 151–158.
[25] D. Culibrk, O. Marques, D. Socek, H. Kalva, B. Furht, Neural network approach to background modeling for video object segmentation, IEEE Trans. Neural Netw. 8 (6) (2007) 1614–1627.
[26] R.M.L. Baena, E. Dominguez, D. Lopez-Rodrıguez, E.J. Palomo, A Neighborhood-Based Competitive Network for Video Segmentation and Object Detection, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 877–886.
[27] R.M.L. Baena, E. Dominguez, D. Lopez-Rodrıguez, E.J. Palomo, A Dipolar Competitive Neural Network for Video Segmentation, Adaptive and Natural Computing Algorithms, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 109–118.
[28] R.M.L. Baena, E.L. Rubio, E. Domínguez, E.J. Palomo, J.M. Jerez, A Self-organizing Map for Traffic Flow Monitoring, Springer, Berlin Heidelberg, 2013, pp. 458–466.
[29] Z.H. Zhou, Y. Jiang, Y.B. Yang, S.F. Chen, Lung cancer cell identification based on artificial neural network ensembles, Artif. Intell. Med. 24 (1) (2002) 25–36.
[30] L.Á. Menéndez, F.J.D.C. Juez, F.S. Lasheras, J.A.Á. Riesgo, Artificial neural networks applied to cancer detection in a breast screening programme, Math. Comput. Modell. 52 (7–8) (2010) 983–991.
[31] L. Maddalena, A. Petrosino. 3D Neural Model-Based Stopped Object Detection, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 585–593.
[32] J.A.R. Quintana, M.I.C. Murguia, An adaptive unsupervised neural network based on perceptual mechanism for dynamic object detection in videos with real scenarios, Neural Process. Lett. (2014) 1–25.
[33] L. Maddalena, A. Petrosino, A self-organizing approach to background subtraction for visual surveillance applications, IEEE Trans. Image Process. 17 (7) (2008) 1168–1177.
[34] G.R. Alonso, M.I.C. Murguia, Object detection in video sequences by a temporal modular self-adaptive SOM, Neural Comput. Appl. (2015) 1–20.
[35] L. Maddalena, A. Petrosino, The3dSOBS+algorithm for moving object detection, Comput. Vision Image Understanding 122 (2014) 65–73.
[36] N. Ikeda, A. Saitoh, T. Isokawa, N. Kamiura, N. Matsui, A neural network approach for counting pedestrians from video sequence images, in: of SICE Annual Conference, Tokyo, 2008, pp. 2485–2488.
[37] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117.
[38] Y. Bengio, Learning deep architectures for AI, Found Trends Mach. Learn. 2 (1) (2009) 1–27.
[39] D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, S. Tubaro, Deep convolutional neural networks for pedestrian detection, Signal Process.-Image Commun. (2016), in press.
[40] X. Jiang, Speed up deep neural network based pedestrian detection by sharing features across multi-scale models, Neurocomputing 185 (2016) 163–170.
[41] N. Najva, E.K. Bijoy, SIFT, tensor based object detection and classification in videos using deep neural networks, Procedia Comput. Sci. 93 (2016) 351–358.
[42] J.G. Rodriguez, S.O. Escolano, A. Angelopoulou, A. Psarrou, J.A. Lopez, J.M.G. Chamizo, Real time motion estimation using a neural architecture implemented on GPUs, J. Real-Time Image 11 (4) (2016) 731–749.
[43] M. Braham, M.V. Droogenbroeck, Deep background subtraction with scene-Specific convolutional neural networks, in: IEEE 23rd International Conference on Systems, Signals and Image Processing, Bratislava, Slovakia, 2016, pp. 1–4.
[44] J.M.O.L. Lobato, R.M. Luque, D.L. Rodríguez, E.J. Palomo, Growing Competitive Network for Tracking Objects in Video Sequences, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 109–118.
[45] V. Rowghani, K.A. Asl, Object tracking by mean shift and radial basis function neural networks, J. Real-Time Image Process. (2015) 1–18.
[46] J. Fan, W. Xu, Y. Wu, Y. Gong, Human tracking using convolutional neural networks, IEEE Trans. Neural Netw. 21 (10) (2010) 1610–1623 (already cited).
[47] R.V. Babu, S. Suresh, A. Makur, Robust Object Tracking with radial basis function networks, International Conference on Acoustics, Speech and Signal Processing, ICASSP 1 (2007) 937–940.
[48] A. Asvadi, M.R.K. Mollaie, Y. Baleghi, Improved object tracking using radial basis function neural networks, in: 7th Iranian Conference on Machine Vision and Image Processing, Tehran, 2011, pp. 1–5.
[49] S. Suresh, F. Brémond, M. Thonnatb, H.J. Kim, Online learning neural tracker, Neurocomputing 74 (10) (2011) 1681–1695.
[50] A. Dimou, P. Medentzidou, F. Alvarez Garcıa, P. Daras, Multi-Target detection in CCTV footage for tracking applications using Deep Learning Techniques, in: IEEE International Conference on Image Processing (ICIP), Phoenix, Arizona, USA, 2016, pp. 928–932.
[51] H.Y. Xue, Y.L.D. Cai, X. He, Tracking people in RGBD videos using deep learning and motion clues, Neurocomputing 204 (2016) 70–76.
[52] G. Wu, W. Lu, G. Gao, C. Zhao, J. Liu, Regional deep learning model for visual tracking, Neurocomputing 175 (2016) 310–323.
[53] Y. Chen, X. Yang, B. Zhong, S. Pan, D. Chen, H. Zhang, CNNTracker Online discriminative object tracking via deep convolutional neural network, Appl. Soft Comput. 38 (2016) 1088–1098.
[54] Y. Chen, X. Yang, B. Zhong, S. Pan, D. Chen, H. Zhang, Network in network based weakly supervised learning for visual tracking, J. Visual Commun. Image Represent. 37 (2016) 3–13.
[55] Tao He, Hua Mao, Moving object recognition using multi-view three-dimensional convolutional neural networks, Neural Comput. Appl. (2016) 1–9.
[56] W. Kim, C. Kim, Background subtraction for dynamic texture scenes using fuzzy color histograms, IEEE Signal Process. Lett. 19 (3) (2012) 127–130.
[57] F.E. Baf, T. Bouwmans, B. Vachon, Type-2 Fuzzy Mixture of Gaussians Model: Application to Background Modeling, Springer-Verlag, Berlin Heidelberg, 2008, pp. 772–781.
[58] J. Zeng, L. Xie, Z.Q. Liu, Type-2 fuzzy Gaussian mixture models, Pattern Recogn. 41 (12) (2008) 3636–3643.
[59] T. Bouwmans, Background subtraction for visual surveillance: a fuzzy approach, in: Handbook on Soft Computing for Video Surveillance, Chapter 5, Taylor and Francis Group, 2012, pp. 103–138.
[60] T. Bouwmans, F.E. Baf, Modeling of dynamic backgrounds by type-2 fuzzy gaussians mixture models, MASAUM J. Basics Appl. Sci. 1 (2) (2009) 265–277.
[61] B. Yao, H. Hagras, D.A. Ghazzawi, M.J. Alhaddad, An Interval Type-2 Fuzzy Logic System for Human Silhouette Extraction in Dynamic Environments, Springer-Verlag, Berlin, Heidelberg, 2012, pp. 126–134.
[62] H. Zhang, D. Xu, Fusing color and texture features for background model, Third International Conference on Fuzzy Systems and Knowledge Discovery (2006) 887–893.
[63] F.E. Baf, T. Bouwmans, B. Vachon, A fuzzy approach for background subtraction, in: International Conference of Image Processing, San Diego, United States, 2008, pp. 2648–2651.
[64] F.E. Baf, T. Bouwmans, B. Vachon, Fuzzy integral for moving object detection, in: International Conference on Fuzzy Systems, IEEE Press New York, 2008, pp. 1729–1736.
[65] Y. Ding, W. Li, J. Fan, H. Yang, A fuzzy background model for moving object detection, 11th IEEE International Conference on Computer-Aided Design and Computer Graphics (2009) 610–613.
[66] M. Sugeno, S.H. Kwon, A new approach to time series modeling with fuzzy measures and the Choquet integral, IEEE International Joint Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium 2 (1995) 799–804.
[67] Y. Narukawa, T. Murofushi, Decision modeling using the Choquet integral, in: Modeling Decisions for Artificial Intelligence, Springer, 2004, pp. 183–193.
[68] M. Balcilar, A.C. Sonmez, Region Based Fuzzy Background Subtraction Using Choquet Integral, Springer-Verlag, Berlin, Heidelberg, 2013, pp. 287–296.
[69] A. Gayathri, A. Srinivasan. Moving Object Detection by Fuzzy Aggregation Using Low Rank Weightage Representation, Springer-Verlag, Berlin, Heidelberg, 2015, pp. 335–342.
[70] D.K. Panda, S. Meher, Detection of moving objects using fuzzy color difference histogram based background subtraction, IEEE Signal Process Lett. 23 (1) (2016) 45–49.
[71] R.T. Iqbal, C. Barbu, F. Petry, Fuzzy component based object detection, Int. J. Approximate Reasoning 45 (2007) 546–563.
[72] R. Iqbal, U. Qidwai, Boosted human-centric hybrid fuzzy classifier, 43rd Annual ACM South East Conference 1 (2005) 36–41.

[73] S.M. balakrishnan, D. Manjula, Fuzzy rule-based classification of human tracking and segmentation using color space conversion, Int. J. Artif. Intell. Appl. 1 (4) (2010) 70–80.
[74] V. Zeljkovic, A. Dorado, E. Izquierdo, Combining a fuzzy rule-based classifier and illumination invariance for improved building detection, IEEE Trans. Circ. Syst. Video Technol. 14 (11) (2004) 1277–1280.
[75] V.F. Arguedas, E. Izquierdo, Object classification based on behaviour patterns, in: Imaging for Crime Detection and Prevention, London, UK, 2011, pp. 1–6.
[76] G. Milanesi, A. Sarti, S. Tubaro, Robust real-Time intrusion detection with fuzzy classification, IEEE International Conference on Image Processing (2002) 437–440.
[77] M.S. Hosseini, A.M.E. Moghadam, Fuzzy rule-based reasoning approach for event detection and annotation of broadcast soccer video, Appl. Soft Comput. 13 (2013) 846–866.
[78] E. Vats, C.S. Chan, Early detection of human actions—A hybrid approach, Appl. Soft Comput. 46 (2016) 953–9661.
[79] J. Ghasemi, R. Ghaderi, M.R. Karami Mollaei, S.A. Hojjatoleslami, A novel fuzzy Dempster-Shafer inference system for brain MRI segmentation, Inf. Sci. 223 (2013) 205–220.
[80] H. Wu, F. Sun, H. Liu, Fuzzy particle filtering for uncertain systems, IEEE Trans. Fuzzy Syst. 16 (5) (2008) 1114–1129.
[81] V. Thomas, A.K. Ray, Fuzzy particle filter for video surveillance, IEEE Trans. Fuzzy Syst. 19 (5) (2011) 937–945.
[82] P.D.Z. Varcheie, G.A. Bilodeau, Adaptive fuzzy particle filter tracker for a PTZ camera in an IP surveillance system, IEEE Trans. Instrum. Meas. (2010) 354–371.
[83] Q. Lin, X.D. Xu, S.T. Wang, Fuzzy particle filter for target tracking, AASRI Procedia (3) (2012) 191–196.
[84] J.M. Garcia, L.R. Benitez, A.F. Caballero, M.T. Lopez, Video sequence motion tracking by fuzzification techniques, Appl. Soft Comput. 10 (2010) 318–331.
[85] E. Parrilla, J.R. Torregrosa, J. Riera, J.L. Hueso, Fuzzy control for obstacle detection in stereo video sequences, Math. Comput. Modell. 54 (2011) 1813–1817.
[86] C. Yoon, M. Cheon, M. Park, Object tracking from image sequences using adaptive models in fuzzy particle filter, Inf. Sci. 253 (2013) 74–99.
[87] R. Paúl, E. Aguirre, M.G. Silvente, R.M. Salinas, A new fuzzy based algorithm for solving stereo vagueness in detecting and tracking people, Int. J. Approximate Reasoning 53 (2012) 693–708.
[88] S.A. Fernandeza, C.A. Lopeza, J.R. Alzolab, A fuzzy-controlled Kalman filter applied to stereo-visual tracking schemes, Signal Process. 83 (2003) 101–120.
[89] S. Aslani, H.M. Nasab, Optical flow based moving object detection and tracking for traffic surveillance, Int. J. Electr. Comput. Energ. Electron. Commun. Eng. 7 (9) (2013) 20–25.
[90] Y. Xina, J. Houa, L. Donga, L. Dingb, A self-Adaptive optical flow method for moving object detection in the video sequences, Optik – Int. J. Light Electron Opt. 125 (19) (2014) 5690–5694.
[91] S. Bhattacharyya, U. Maulik, P. Dutta, High-speed target tracking by fuzzy hostility-induced segmentation of optical flow field, Appl. Soft Comput. 9 (2009) 126–134.
[92] B.G. Kim, D.J. Park, Novel target segmentation and tracking based on fuzzy membership distribution for vision-based target tracking system, Image Vision Comput. 24 (2006) 1319–1331.
[93] T. Frantti, Multiphase transfer of control signal for adaptive power control in CDMA systems, Control Engg. Prac. 14 (5) (2006) 489–501.
[94] T. Frantti, P. Mahonen, Fuzzy logic based forecasting model, Eng. Appl. Artif. Intell. 14 (1) (2001) 189–201.
[95] J.H. Holland, Genetic Algorithms, 2018, http://www2.econ.iastate.edu/tesfatsi/holland.GAIntro.htm. (Accessed 28 September 2016).
[96] F. Marini, B. Walczak, Particle swarm optimization (PSO): a tutorial, Chemom. Intell. Lab. Syst. 149 (B) (2015) 153–165.
[97] L. Ingber, Simulated annealing: practice versus theory, Math. Comput. Modell. 18 (11) (1993) 29–57.
[98] B. Karasulu, S. Korukoglu, Moving object detection and tracking by using annealed background subtraction method in videos: performance optimization, Expert Syst. Appl. 39 (2012) 33–43.
[99] E.Y. Kim, K.J. Sang, W. Hwang, H.J. Kim, Video sequence segmentation using genetic algorithms, Pattern Recognit. Lett. 23 (7) (2002) 843–863.
[100] E.Y. Kim, K. Jung, Object Detection and Removal Using Genetic Algorithms, Springer-Verlag, Berlin, Heidelberg, 2004, pp. 411–421.
[101] E.Y. Kima, S.H. Park, Automatic video segmentation using genetic algorithms, Pattern Recognit. Lett. 27 (11) (2006) 1252–1265.
[102] T. Kim, S. Lee, J. Paik, Evolutionary Algorithm-Based Background Generation for Robust Object Detection, Springer-Verlag, Berlin, Heidelberg, 2006, pp. 542–552.
[103] J. Yu, X. Zhou, F. Qian, A novel approach for moving object detection based on improved particle swarm optimization algorithm, in: 8th IEEE International Conference on Control and Automation, China, 2010, pp. 1178–1183.
[104] S. Cagnoni, M. Mordonini, J. Sartori, Particle Swarm Optimization for Object Detection and Segmentation, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 241–250.
[105] P. Bhuvaneswari, A.B. Therese, Detection of cancer in lung with K-NN classification using genetic algorithm, Procedia Mater. Sci. 10 (2015) 433–440.
[106] G.R. Chandra, K.R.H. Rao, Tumor detection In brain using genetic algorithm, Procedia Comput. Sci. 79 (2016) 449–457.
[107] D.Y. Liu, H.L. Chen, B. Yang, X.E. Lv, L.N. Li, J. Liu, Design of an enhanced fuzzy k-nearest neighbor classifier based computer aided diagnostic system for thyroid disease, J. Med. Syst. 36 (5) (2012) 3243–3254.
[108] S.L. Glioma, Detection based on multi-fractal features of segmented brain MRI by particle swarm optimization techniques, Biomed. Signal Process. Control 31 (2017) 148–155.
[109] M.E. Sweety, G.W. Jiji, Detection of Alzheimer disease in brain images using PSO and Decision Tree Approach, in: International Conference on Advanced Communication Control and Computing Technologies, India, 2014, pp. 1305–1309.
[110] H.S. Boroujeni, N.M. Charkari, M. Behrouzifar, P.T. Makhsoos, Tracking Multiple Variable-Sizes Moving Objects in LFR Videos Using a Novel Genetic Algorithm Approach, Springer-Verlag, Berlin, Heidelberg, 2012, pp. 143–153.
[111] M. Speekenbrink, A tutorial on particle filters, J. Math. Psychol. 73 (2016) 140–152.
[112] G. Oppenheima, A. Philippe, J. de Rigal, The particle filters and their applications, Chemom. Intell. Lab. Syst. 91 (1) (2008) 87–93.
[113] H. Han, Y.S. Ding, K.R. Hao, X. Liang, An evolutionary particle filter with the immune genetic algorithm for intelligent video target tracking, Comput. Math. Appl. 62 (7) (2011) 2685–2695.
[114] M. Thidaa, H.L. Eng, D.N. Monekossob, P. Remagnino, A particle swarm optimisation algorithm with interactive swarms for tracking multiple targets, Appl. Soft Comput. 13 (6) (2013) 3106–3117.
[115] X. Zhang, W. Hu, S. Maybank, X. Li, M. Zhu, Sequential particle swarm optimization for visual tracking, IEEE International Conference on Computer Vision and Pattern Recognition (2008) 1–8.
[116] L.A. Canalis, M.H. Tejera, E.S. Nielsen, Swarm track: a particle swarm approach to visual tracking, Portugal, in: International Conference on Computer Vision Theory and Applications, 2, 2006, pp. 221–228.
[117] Y. Zhang, Y. Meng, Adaptive object tracking using particle swarm optimization, IEEE International Symposium on Computational Intelligence in Robotics and Automation (2007) 43–48.
[118] S. Yang, Q. Ma, W. Huang, Particle swarm optimized unscented particle filter for target tracking, International Congress on Image and Signal Processing (2009) 1–5.
[119] H. Zhou, X. Zhang, H. Li, J. Li, Video object tracking based on swarm optimized particle filter, International Conference on Industrial Mechatronics and Automation 2 (2010) 702–706.
[120] C. Bae, K. Kang, G. Liu, Y.Y. Chung, A novel real time video tracking framework using adaptive discrete swarm optimization, Expert Syst. Appl. 64 (2016) 385–399.
[121] Z. Wang, X. Wu, Salient object detection using biogeography-based optimization to combine features, Appl. Intell. 45 (1) (2016) 1–17.
[122] M. Zhang, M. Xin, J. Yang, Adaptive multi-cue based particle swarm optimization guided particle filter tracking in infrared videos, Neurocomputing 25 (2013) 163–171.
[123] A.M.A. Tawaba, M.B. Abdelhalimb, S.E.D. Habib, Efficient multi-feature PSO for fast gray level object-tracking, Appl. Soft Comput. 14 (2014) 317–337.
[124] R.F.A. Kader, R. Atta, S.E. Shakhabe, An efficient eye detection and tracking system based on particle swarm optimization and adaptive block-matching search algorithm, Eng. Appl. Artif. Intell. 31 (2014) 90–100.
[125] J. Zhao, Z. Li, Particle filter based on Particle Swarm Optimization resampling for vision tracking, Expert Syst. Appl. 37 (2010) 8910–8914.
[126] L. Maddalena, A. Petrosino, Multivalued Background/Foreground Separation for Moving Object Detection, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 263–270.
[127] M.I.C. Murguia, S.G. Duarte, An adaptive neural-fuzzy approach for object detection in dynamic backgrounds for surveillance systems, IEEE Trans. Ind. Electron. 59 (8) (2012) 3286–3298.
[128] L. Maddalena, A. Petrosino, A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection, Neural Comput. Appl. 19 (2010) 179–186.
[129] R.M. Luque, D.L. Rodrıguez, E.M. Casermeiro, Video object segmentation with multivalued neural networks, IEEE Eighth International Conference on Hybrid Intelligent Systems (2008) 613–618.
[130] C. Stauffer, W. Grimson, Adaptive background mixture models for real-time tracking, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1999).
[131] E.C. Gallego, P. Brox, S.S. Solano, A Fuzzy System for Background Modeling in Video Sequences, Springer-Verlag, Berlin, Heidelberg, 2013, pp. 184–192.
[132] M.I.C. Murguia, G.R. Alonso, Fuzzy-neural self-adapting background modeling with automatic motion analysis for dynamic object detection, Appl. Soft Comput. 36 (2015) 570–577.
[133] Z. Li, W. Liu, Y. Zhang, Adaptive fuzzy approach to background modeling using PSO and KLMs, 10th World Congress on Intelligent Control and Automation at Bejing (2012) 4601–4607.
[134] E. PaulIjjina, K.M. Chalavadi, Human action recognition using genetic algorithms and convolutional neural networks, Pattern Recogn. 59 (2016) 199–212.
[135] E.P. Ijjina, C.K. Mohan, Hybrid deep neural network model for human action recognition, Appl. Soft Comput. 46 (2016) 936–952.
[136] C.F. Wu, C.J. Lin, H.Y. Lin, H. Chung, Adjacent lane detection and lateral vehicle distance measurement using vision-based neuro-fuzzy approaches, J. Appl. Res. Technol. 11 (2) (2013) 251–258.

[137] G. Acampora, P. Foggia, A. Saggese, M. Vento, A hierarchical neuro-fuzzy architecture for human behavior analysis, Inf. Sci. 310 (2015) 130–148.
[138] S.Y. Cho, C. Quek, S.X. Seah, C.H. Chong, HebbR2-Taffic. A novel application of neuro-fuzzy network for visual based traffic monitoring system, Expert Syst. Appl. 36 (3 Part 2) (2009) 6343–6356.
[139] S. Cho, B.Y. Kim, E.H. Park, Y. Chang, J. Kim, K. Chung, W. Whang, H. Kim, Automatic Recognition of Alzheimer's Disease Using Genetic Algorithms and Neural Network, Springer-Verlag, Berlin, Heidelberg, 2003, pp. 695–702.
[140] H.T. Kim, B.Y. Kim, E.H. Park, J.W. Kim, E.W. Hwang, S.K. Han, S. Cho, Computerized recognition of Alzheimer disease-EEG using genetic algorithms and neural network, Future Generat. Comput. Syst. 21 (7) (2005) 1124–1130.
[141] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, T.F. Yuan, Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning, Front. Comput. Neurosci. 9 (66) (2015) 1–15.
[142] Y. Zhang, S. Wang, Z. Dong, P. Phillip, G. Ji, J. Yang, Pathological brain detection In magnetic resonance imaging scanning by wavelet entropy and hybridization of biogeography-based optimization and particle swarm optimization, Progress Electromagnet. Res. 152 (2015) 41–58.
[143] S. Wang, Y. Zhang, Z. Dong, S. Du, G. Ji, J. Yan, J. Yang, Q. Wang, C. Feng, P. Phillips, Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection, Int. J. Imaging Syst. Technol. 25 (2) (2016) 153–164.
[144] Y. Zhang, S. Wang, P. Phillips, Z. Dong, G. Ji, J. Yang, Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC, Biomed. Signal Process. Control 21 (2015) 58–73.
[145] C. Gao, P. Li, Y. Zhang, J. Liu, L. Wang, People counting based on head detection combining Adaboost and CNN in crowded surveillance environment, Neurocomputing 208 (2016) 108–116.
[146] W.C. Cheng, D.M. Jhan, A self-constructing cascade classifier with AdaBoost and SVM for pedestrian detection, Eng. Appl. Artif. Intell. 26 (3) (2013) 1016–1028.
[147] L. Guo, P.S. Ge, M.H. Zhang, L.H. Li, Y.B. Zhao, Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine, Expert Syst. Appl. 39 (4) (2012) 4274–4286.
[148] B. Scholkopf, A.J. Smola, Learning With Kernels, MIT Press, Cambridge, MA, 2002.
[149] S. Zhang, Y. Sui, X. Yu, S. Zhao, L. Zhang, Hybrid support vector machines for robust object tracking, Pattern Recogn. 48 (8) (2015) 2474–2488.
[150] B. Cyganek, Framework for object tracking with support vector machines, structural tensor and the mean shift method, in: Neural Information Processing, Springer, Berlin, Heidelberg, 2009, pp. 399–408.
[151] B. Cyganek, M. Woźniak, Pixel-Based object detection and tracking with ensemble of support vector machines and extended structural tensor, in: Computational Collective Intelligence. Technologies and Applications, Springer, Berlin, Heidelberg, 2012, pp. 104–113.
[152] S. Zhang, X. Yu, Y. Sui, S. Zhao, L. Zhang, Object tracking with multi-view support vector machines, IEEE Trans. Multimedia 17 (3) (2015) 265–278.
[153] Y. Bengio, A. Courville, P. Vincent, Representation Learning: A Review and New Perspectives. Technical Report, U. Montreal, 2012 URL http://arxiv.org/abs/1206.5538.
[154] Y. Zhang, X. Li, Z. Zhang, F. Wu, Liming Zhao, Deep learning driven blockwise moving object detection with binary scene modeling, Neurocomputing 168 (2015) 454–463.
[155] S.Z. Su, Z.H. Liu, S.P. Xu, S.Z. Li, R. Ji, Sparse auto-encoder based feature learning for human body detection in depth image, J. Signal Process. 112 (C) (2015) 43–52.
[156] P. Ji, L. Cao, X. Zhang, L. Zhang, W. Wu, News videos anchor person detection by shot clustering, Neurocomputing 123 (2014) 86–99.
[157] R. Yao, Y. Zhangb, Y. Zhoua, S. Xiaa, Multiple small objects tracking based on dynamic Bayesian networks with spatial prior, Optik – Int. J. Light Electron Opt. 125 (10) (2014) 2243–2247.
[158] P. Premaratne, S. Yang, P. Vial, Z. Ifthikar, Centroid tracking based dynamic hand gesture recognition using discrete Hidden Markov Models, Neurocomputing, (Available online 29 October 2016).
[159] B. Wua, B.G. Hua, Q. Jic, A Coupled Hidden Markov Random Field Model for Simultaneous Face Clustering and Tracking in Videos, Pattern Recognition, (Available online 26 October 2016).
[160] T. Vojir, J. Matasa, J. Noskovab, Online adaptive hidden Markov model for multi-tracker fusion, Computer Vision and Image Understanding, (Available online 16 May 2016).
[161] L.R. Benitez, C.S. Cipres, J.M. Garcia, L.J. Linares, Approximate reasoning and finite state machines to the detection of actions in video sequences, Int. J. Approximate Reasoning 52 (2011) 526–540.
[162] B. Tian, M. Tang, F.Y. Wang, Vehicle detection grammars with partial occlusion handling for traffic surveillance, Transp. Res. Part C: Emerg. Technol. 56 (2015) 80–93.
[163] P. Jackson, Introduction To Expert Systems, 3 ed., Addison Wesley, 1998, p. 2, ISBN 978-0-201-87686-4.
[164] D. Alvarez, R. Fernández, L. Sánchez, J. Alija, Expert system for handwritten numeral recognition using dynamic zoning, International Conference on Hybrid Artificial Intelligence Systems (2015) 125–135.
[165] E. Yaya, N.M. Madrida, J.A.O. Ramírez, Using an improved rule match algorithm in an expert system to detect broken driving rules for an energy-efficiency and safety relevant driving system, Procedia Comput. Sci. 35 (2014) 127–136.
[166] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, Wallflower: principles and practice of background maintenance, in: Seventh International Conference on Computer Vision, Kerkyra, Greece, 1999, pp. 255–261.
[167] CAVIAR, Context Aware Vision Using Image-Based Active Recognition, 2002 http://homepages.inf.ed.ac.uk/rbf/CAVIAR/.
[168] Advanced video and signal based surveillance 2007, datasets, December 2011. http://www.eecs.qmul.ac.uk/andrea/avss2007d.html.
[169] L. Gorelick, M. Blank, E. Shechtman, M. Irani, R. Basri, Actions as space-time shapes, IEEE Trans. Pattern Anal. Mach. Intell. 29 (12) (2007) 2247–2253.
[170] C. Sanderson, B. Lovell, Multi-region probabilistic histograms for robust and scalable identity inference, Lect. Notes Comput. Sci. 5558 (2009) 199–208.
[171] K.K. Reddy, M. Shah, Recognizing 50 human action categories of web videos, Mach. Vision. Appl. 24 (5) (2013) 971–981.
[172] Y. Wu, J.W. Lim, M.H. Yang, Online object tracking: a benchmark, in: IEEE Conference on Computer Vision and Pattern Recognition, Portland, 2013, pp. 2411–2418.
[173] G. Fanelli, T. Weise, J. Gall, L. Van Gool, Real Time Head Pose Estimation from Consumer Depth Cameras, 33rd Annual Symposium of the German Association for Pattern Recognition (DAGM'11).
[174] INRIA, Etiseo video understanding evaluation, December 2011. http://www-sop.inria.fr/orion/ETISEO/index.htm.
[175] Tracking Revisited using RGBD Camera: Unified Benchmark and Baselines Proceedings of 14th IEEE International Conference on Computer Vision (ICCV2013).
[176] P. Liang, E. Blasch, Haibin Ling, Encoding color information for visual tracking: algorithms and benchmark, IEEE Trans. Image Process. 24 (12) (2015) 5630–5644.
[177] Oklahoma State University, Otcbvs benchmark dataset collection, November 2011. http://www.cse.ohio-state.edu/otcbvs-bench/.
[178] Background Models Challenge (BMC) dataset: A. Vacavant, T., Chateau, A. Wilhelm, L. Lequievre, A Benchmark Dataset for Foreground/Background Extraction, Computer Vision – ACCV 2012 Workshops, Springer Berlin Heidelberg, Vol. 7728, ISBN: 978-3-642-37410-4, 2012, pp. 291–300. http://bmc.iut-auvergne.com.
[179] N. Goyette, P.M. Jodoin, F. Porikli, J. Konrad, P. Ishwar, Changedetection.net: a new change detection benchmark dataset, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (2012) 1–8.
[180] Tenth IEEE international workshop on performance evaluation of tracking and surveillance, Jan uary 2012. http://pets2007.net/.
[181] L. Li, T. Nawaz, J. Ferryman, PETS 2015: datasets and challenge, in: 12th IEEE International Conference on Advanced Video and Signal Based Surveillance, Karlsruhe, 2015, pp. 1–6.
[182] A. Sanin, Conrad Sanderson, Brian C. Lovell, Shadow detection: a survey and comparative evaluation of recent methods, Pattern Recogn. 45 (2012) 1684–1695.
[183] N.A.N. Helmut, E. Bez, J. Singhai, E.A. Edirisinghe, A survey of cast shadow detection algorithms, Pattern Recognit. Lett. 33 (6) (2012) 752–764.
[184] B. Javidya, A. Hatamloua, S. Mirjalilib, Ions motion algorithm for solving optimization problems, Appl. Soft Comput. 32 (2015) 72–79.
[185] A. Sadollaha, H. Eskandarb, J.H. Kima, Water cycle algorithm for solving constrained multi-objective optimization problems, Appl. Soft Comput. 27 (2015) 279–298.
[186] Y.J. Zheng, Water wave optimization: a new nature-inspired metaheuristic, Comput. Oper. Res. 55 (2015) 1–11.
[187] A. Sadollaha, A. Bahreininejada, H. Eskandarb, M. Hamdia, Mine blast algorithm: a new population based algorithm for solving constrained engineering optimization problems, Appl. Soft Comput. 13 (5) (2013) 2519–2612.
[188] P. Shanmugavadivu, K. Balasubranium, Particle swarm optimized multi-objective histogram equalization for image enhancement, Opt. Laser Technol. 57 (2014) 243–251.
[189] Y. Wang, H. Wang, C. Yin, M. Dai, Biologically inspired image enhancement based on Retinex, Neurocomputing 177 (2016) 373–384.
[190] Jun Lei, Guo Hui Li, Dan Tu, Qiang Guo, Convolutional restricted Boltzmann machines learning for robust visual tracking, Neural Comput. Appl. 25 (November (6)) (2014) 1383–1391.
[191] M. Havaei, A. Davy, D.W. Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.M. Jodoin, H. Larochelle, Brain tumor segmentation with deep neural networks, Med. Image Anal. 35 (2017) 18–31.
[192] F. Valdez, P. Melin, O. Castillo, Modular Neural Networks architecture optimization with a new nature inspired method using a fuzzy combination of Particle Swarm Optimization and Genetic Algorithms, Inf. Sci. 270 (2014) 143–153.
[193] H. Yang, L. Shao, F. Zheng, L. Wang, Z. Song, Recent advances and trends in visual tracking: a review, Neurocomputing 74 (18) (2011) 3823–3831.
[194] Naiyan Wang, Dit-Yan Yeung, Learning a deep compact image representation for visual tracking, Adv. Neural Inf. Process. Syst. 26 (2013) 1–13.
[195] Seunghoon Hong, Tackgeun Suha Kwak, Bohyung Han, Online tracking by learning discriminative saliency map with convolutional neural network, in: 32nd International Conference on Machine Learning, Lille, France, 2015, p. 37.

[196] Le Zhang, Ponnuthurai Nagaratnam Suganthan, Visual tracking with convolutional neural network, in: IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, 2015, pp. 2072–2077.
[197] Le Zhang, Ponnuthurai Nagaratnam Suganthan, Visual tracking with convolutional random vector functional link network, IEEE Trans. Cybern. (2016) 1–11.
[198] Lijun Wang, Wanli Ouyang, Xiaogang Wang, Huchuan Lu, Visual tracking with fully convolutional networks, IEEE International Conference on Computer Vision (2015) 1–9.
[199] Anton Milan, Seyed Hamid Rezatofighi, Anthony Dick, Ian Reid, Konrad Schindler, Online multi-target tracking using recurrent neural networks, Comput. Vis. Pattern Recognit. (2016).
[200] Z. Kalal, K. Mikolajczyk, J. Matas, Tracking-learning-detection, IEEE Trans. Pattern Anal. Mach. Intell. 6 (1) (2010).
[201] M.D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, L. Van Gool, Robust tracking-by-Detection using a detector confidence particle filter, IEEE 12th International Conference on Computer Vision (2009) 2009.
[202] M. Andriluka, S. Roth, B. Schiele, People-Tracking-by-Detection and people-Detection-by-Tracking. people-Tracking-by-Detection and people-Detection-by-Tracking, in: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'08), Anchorage, USA, June, 2008.
[203] Eunae Park, Hyuntae Ju, Yong Mu Jeong, Soo-Young Min, Tracking-Learning-Detection adopted unsupervised learning algorithm, IEEE Seventh International Conference on Knowledge and Systems Engineering (KSE) (2015).
[204] Severin Stalder, Helmut Grabner, Luc Van Gool, Cascaded Confidence Filtering for Improved Tracking-by-Detection, Springer, Berlin, Heidelberg, 2010, pp. 369–382.
[205] Avidan Shai, Ensemble tracking, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2) (2007) 261–271.
[206] Q. Bai, Z. Wu, S. Sclaroff, M. Betke, C. Monnier, Randomized ensemble tracking, Proceedings of the IEEE International Conference on Computer Vision (2013).
[207] Lior Rokach, Ensemble-based classifiers, Artif. Intell. Rev. 33 (2010) 1.
[208] Ye Ren, Le Zhang, P.N. Suganthan, Ensemble classification and regression-recent developments, applications and future directions [Review article], IEEE Comput. Intell. Mag. 11 (1) (2016) 41–53.
[209] D.S. Bolme, B.A. Draper, J.R. Beveridge, Average of synthetic exact filters, Cvpr (2009).
[210] David S. Bolme, J. Ross Beveridge, Bruce A. Draper, Yui Man Lui, Visual Object Tracking Using Adaptive Correlation Filters, 2018 http://www.cs.colostate.edu/~draper/papers/bolme_cvpr10.pdf.
[211] K.-H. Jeong, P.P. Pokharel, J.-W. Xu, S. Han, J. Principe, Kernel based synthetic discriminant function for object recognition, ICASSP (2006).
[212] C. Xie, M. Savvides, B. Vijaya-Kumar, Kernel correlation filter based redundant class-dependence feature analysis (KCFA) on FRGC 2.0 data, in: Analysis and Modeling of Faces and Gestures, 2005.
[213] Yi Wu, Jongwoo Lim, Ming-Hsuan Yang, Object tracking benchmark, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1834–1848.
[214] A. Milan, L. Leal-Taixe, Ian Reid, S. Roth, Konrad Schindler, MOTChallenge: The Multiple Object Tracking Benchmark, 2018, Available at: https://motchallenge.net/. (Accessed 14 April 2017).
[215] A. Li, M. Lin, Y. Wu, M.H. Yang, S. Yan, Nus-pro: a new visual tracking challenge, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2016) 335–349.
[216] M. Kristan, J. Matas, A. Leonardis, T. Voj, R. Pflugfelder, G. Fernandez, G. Nebehay, F. Porikli, L. Cehovin, A novel performance evaluation methodology for single-target trackers, IEEE Trans. Pattern Anal. Mach. Learn. (2016) (Accepted). online repository: http://votchallenge.net/howto/. (Accessed 14 April 2017).
[217] M. Kaushal, B. Singh, Akashdeep, Performance evaluation of fuzzy 2-partition entropy and big bang big crunch optimization based object detection and tracking approach, Multidimension. Syst. Signal Process. (2017), http://dx.doi.org/10.1007/s11045-017-0515-7 (Online).
[218] Yuanhao Yu, Qingsong Wu, T. Kirubarajan, Yasuo Uehara, Robust discriminative tracking via structured prior regularization, Image Comput. Vis. 69 (2018).
[219] L. Liang-qun, Z. Xi-yang, L. Zong-xiang, X. Wei-xin, Fuzzy logic approach to visual multi-object tracking, Neuro Computing, (Available online 5 December 2017).